



Automates cellulaires pour la modélisation multi-échelle des systèmes biologiques

Benjamin Louvet

► To cite this version:

Benjamin Louvet. Automates cellulaires pour la modélisation multi-échelle des systèmes biologiques. Autre [cs.OH]. Université Blaise Pascal - Clermont-Ferrand II, 2014. Français. NNT : 2014CLF22479 . tel-01276668

HAL Id: tel-01276668

<https://theses.hal.science/tel-01276668>

Submitted on 19 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE CLERMONT FERRAND II – BLAISE PASCAL

ÉCOLE DOCTORALE SCIENCES POUR L'INGENIEUR

Thèse de Doctorat, Spécialité Informatique

Présentée par

Benjamin Louvet

Pour obtenir le grade de

DOCTEUR D'UNIVERSITE, SPECIALITE INFORMATIQUE

AUTOMATES CELLULAIRES POUR LA MODELISATION MULTI-ECHELLE DES SYSTEMES BIOLOGIQUES

Soutenue publiquement le 11 Juillet 2014 devant le jury :

M. George Czaplicki	Professeur des Universités	Rapporteur
M. Laurent Deroussi	Maître de Conférences	Co-directeur de Thèse
M. Michel Gourgand	Professeur des Universités	Directeur de Thèse
M. Samir Lamouri	Professeur des Universités	Examineur ; Président du jury
M. David Sarramia	Maître de Conférences	Co-directeur de Thèse
M. Patrick Siarry	Professeur des Universités	Rapporteur

RESUME ET MOTS CLES

RESUME

Ce projet de thèse, dans le cadre d'une collaboration entre le LIMOS et le LPC, s'inscrit dans une démarche de recherche permettant la mise en synergie des domaines de la biologie, de la physique et de l'informatique par la proposition d'une démarche de simulation permettant la réalisation d'expériences *in silico*.

Pour cela, nous nous proposons de développer une plateforme logicielle dédiée à la modélisation multiéchelle des systèmes biologiques qui pourra par la suite être interfacée avec les outils de simulation de physique des particules. Nous proposons également un modèle individu-centré de cellule biologique paramétrable à l'aide de données obtenues d'expériences *in vitro*.

Nous présentons l'élaboration de cette plateforme et une démarche de validation de ses fonctionnalités à travers l'implémentation de modèles d'automates cellulaires de la littérature. Nous présentons ensuite la construction du modèle de cellule biologique en prenant le temps d'expliquer comment est pris en compte le système biologique, comment nous le modélisons puis comment nous paramétrons le modèle. Nous modélisons les processus internes de la cellule, dont les caractéristiques sont liées à l'information génétique qu'elle porte. Ce modèle de cellule permet de reproduire le comportement d'une cellule isolée, et à partir de là, d'un ensemble de cellules via l'automate. Le modèle est ensuite utilisé pour retrouver les courbes de croissance d'une population de bactéries *Escherichia coli*. Des valeurs de données de fluxomique ont été exploitées et ont permis la reproduction *in silico* des expériences *in vitro* dont elles étaient issues.

MOTS CLES

Automates cellulaires ; Modélisation ; Multi-échelle ; Système biologique ; Biologie intégrative ; Biologie des systèmes ; Cellule ; Cinétique de croissance cellulaire ; *Escherichia coli* ; *In silico* ;

ABSTRACT AND KEYWORDS

ABSTRACT

This PhD thesis project is part of a research program in the fields of biology, physics and computer science aiming to propose a simulation approach for performing experiments *in silico*.

For this, we propose to develop a software platform dedicated to multi-scale modeling of biological systems that can be combined with particle physics simulation tools. We also propose a general individual-based model of biological cell in which data obtained from *in vitro* experiments can be used.

We present the development of this platform and the validation process of its functionalities through the implementation of cellular automata from the literature. We then present the design of the biological cell model by giving the hypothesis we made, how we model and how we parameterize the model. Starting from a simple biological system, bacteria, observed in liquid culture, our model uses a multi-scale middle-out approach. We focus on the cell and we model internal processes, assuming that all their properties come from genetic information carried out by the cell's genome. This model allows to consider the cell behavior, and then to obtain the behavior of a cell population. Data from fluxomic experiments have been used in this model to parameterize the biochemical processes. The results we obtain allow us to consider the model as validated as simulation results match the experimental data.

KEYWORDS

Cellular automata ; Modeling ; Multi-scale ; Biological system ; Integrative biology ; System biology ; Cell ; Cell growth kinetic ; *Escherichia coli* ; *In silico*

UNITES D'ACCUEIL

LIMOS



**Laboratoire d'Informatique, de Modélisation et
d'Optimisation des Systèmes, LIMOS**

UMR 6158
Université Blaise Pascal,
Complexe Scientifique des Cézeaux
63173 Aubière Cedex

LPC



Laboratoire de Physique Corpusculaire, LPC

UMR 6533
Université Blaise Pascal,
Campus des Cézeaux, Bâtiment 7
Avenue Blaise Pascal
BP 80026
63171 Aubière Cedex

IUT D'ALLIER, SITE MONTLUÇONNAIS



IUT d'Allier, site Montluçon

Université Blaise Pascal
Avenue Aristide Briand
CS 82235
03101 Montluçon Cedex

REMERCIEMENTS

Il est traditionnel de trouver dans ce genre de document une rubrique dans laquelle sont remerciées toutes les personnes et entités impliquées de près ou de loin dans l'aboutissement de tels travaux. Avec les événements et les difficultés auxquelles j'ai dû faire face au cours de ces dernières années, elle m'est particulièrement chère.

De façon traditionnelle, je remercie le Conseil Général de l'Allier pour le financement dont j'ai bénéficié pour les trois premières années. De la même manière, je remercie le LIMOS et le LPC pour l'accueil dont j'ai bénéficié et pour le matériel mis à ma disposition pour mener à bien ces travaux. Je ne saurai être exhaustif en citant les noms des membres des équipes impliquées, alors c'est un merci collectif que j'adresse. Les mêmes remerciements sont adressés à l'IUT d'Allier, site de Montluçon et son personnel pour les mêmes raisons, l'accueil et l'aide matérielle.

Six mois d'hospitalisation en fin de thèse, c'est un coup rude. Après pareil arrêt, se réapproprier son sujet, son travail, se remettre en selle, reprendre le rythme, c'est chose difficile, et je le confesse, souvent ça m'a paru insurmontable. C'est dans ce contexte que les remerciements que j'ai à formuler prennent à mes yeux cette valeur toute particulière. Le nous employé dans le présent document n'est pas qu'une formulation académique ; il reflète l'importance de tout cet entourage.

Merci à l'équipe soignante qui m'a soutenu et accompagné, un grand Merci Gérard d'avoir toujours prêté oreille attentive à mon ressenti et d'avoir été disponible pour essayer de concilier gestion des douleurs et capacité de travail.

Merci à Béa au LIMOS qui m'a aidé à me remettre en selle et qui a su me sortir une grosse épine du pied lors de mon retour mouvementé. Merci à Evelyne et à Cécile à l'IUT de m'avoir aidé à faire face à certaines difficultés administratives lors de cette période. Merci à Mme Marchand de m'avoir accueilli dans l'équipe pédagogique de son établissement pour assumer cette dernière année délicate. Merci à cette équipe pédagogique pour les bouffées d'air hors contexte universitaire.

De manière générale, pour leur soutien au cours de ces quelques années, et particulièrement la dernière, j'adresse un grand merci à tous les membres de l'équipe GLT, Anne-Céline, Catherine, Christine, Émilie, Nathalie, Sandra, Sophie, Valérie et Yves-Jean. Un grand merci aussi « aux autres », Alexis, Bruce, Didier, Jérôme, Marc, Myriam, que j'ai eu tant plaisir à côtoyer à l'IUT.

Une pensée particulière pour mes compagnons de galère, les autres doctorants, Fabrice, Paul, Virginie et tout récemment Nathanaël. Mention spéciale pour Fazia, dont la spontanéité, la maladresse et le sens de l'orientation (!!!) auront été source de fou-rires partagés. Je souhaite le meilleur à tous.

Une pensée affectueuse à la famille, dans son ensemble pour son soutien inconditionnel. Même pensée affectueuse aux amis, Amélie, Aurore, Caroline, Christophe, Emili, Méritxell, Sylvain dont la présence a été salvatrice.

Pour finir, parce que sans eux rien de tout ceci n'aurait été fait, ou tout du moins, ce ne serait pas mon nom sur ce manuscrit, merci à David, Laurent et Michel. Merci de m'avoir donné ma chance, de m'avoir accordé votre confiance, de m'avoir soutenu et accompagné au cours de ces années. Et enfin, merci à M. Siarry et M. Czaplicki d'avoir accepté la charge de rapporteur de ces travaux, merci à M. Lamouri d'avoir accepté les rôles d'examineur et de président du jury.

SOMMAIRE

Résumé et mots clés	3
Abstract and keywords	4
Unités d'accueil	5
Remerciements	7
Sommaire	9
Chapitre I. Introduction générale : contexte et problématique	11
I.1. Contexte	13
I.2. Problématique	15
I.3. Plan de la thèse	18
Chapitre II. Propriétés et mécanismes des systèmes biologiques	21
II.1. Introduction	23
II.2. Généralités	24
II.3. L'hérédité et la transmission de patrimoine génétique	29
II.3.1. Principes de l'hérédité	29
II.3.2. Notions de génomique	32
II.4. L'unité fonctionnelle du vivant : la cellule	38
II.4.1. Structure des cellules	38
II.4.2. Vue d'ensemble de la vie des cellules	43
II.4.3. Vue d'ensemble du fonctionnement cellulaire	45
II.5. Environnement	50
II.6. Notions d'adaptation et d'évolution	51
II.6.1. Adaptation	52
II.6.2. Evolution	52
II.7. Cadre expérimental	54
II.7.1. Environnements expérimentaux, notion de culture cellulaire	55
II.7.2. Radiobiologie	59
II.8. Conclusion	62
Chapitre III. Etat de l'art : des modèles aux plateformes	63
III.1. Introduction	65
III.2. Modèles de population cellulaire : différentes approches	67
III.2.1. Modélisation à population pour les processus biologiques	69
III.2.2. Modélisation individu-centrée pour les systèmes biologiques	73
III.2.3. Bilan sur les types de modèles	82
III.3. Modèles pour les processus biologiques	83
III.3.1. Processus biochimiques	86
III.3.2. Adaptation	91
III.3.3. Evolution	96
III.3.4. Synthèse de ce tour d'horizon des modèles	102
III.4. Plateformes	103
III.4.1. Formalismes	103
III.4.2. Plateformes dédiées	104
III.4.3. Plateformes génériques	105
III.5. Conclusion	107
Chapitre IV. Propositions : une plateforme, un modèle de cellule biologique	111

IV.1. Plateforme	113
IV.1.1. Présentation	113
IV.1.2. Applications : modèles et validation de la plateforme	116
IV.1.3. Remarques sur les temps de calculs.....	138
IV.1.4. Conclusion	140
IV.2. Modèle de cellule	141
IV.2.1. Hypothèses de modélisation.....	142
IV.2.2. Paramètres du modèle.....	146
IV.2.3. Devenir des cellules filles.....	150
IV.2.4. Superindividu.....	150
IV.3. Définition des valeurs des paramètres	152
IV.3.1. Consommation par bactérie	153
IV.3.2. Environnement modélisé.....	153
IV.3.3. Densité de population bactérienne donnant une DO de 1.....	154
IV.4. Fonctionnement et implémentation.....	155
IV.4.1. Paramètres de l'automate cellulaire	155
IV.4.2. Gestion et suivi de la simulation.....	155
IV.4.3. Initialisation d'une cellule	157
IV.4.4. Fonctions cellulaires.....	158
IV.5. Conclusion.....	161
Chapitre V. Mise en œuvre du modèle et de la plateforme : résultats	163
V.1. Matériel et méthodes.....	165
V.2. Précision sur le fonctionnement de l'automate cellulaire.....	167
V.3. Reproductibilité.....	168
V.4. Validation du comportement du modèle et de la robustesse des résultats	169
V.4.1. Jeux d'essais	169
V.4.2. Bilan des résultats sans exploiter le superindividu	175
V.4.3. Résultats avec utilisation du superindividu	176
V.5. Utilisation des paramètres issus d'études biologiques	178
V.5.1. Implémentation.....	178
V.5.2. Résultats.....	178
V.5.3. Phase stationnaire.....	180
V.6. Conclusion	183
Chapitre VI. Conclusions et perspectives	185
Bibliographie.....	189
Glossaire.....	199

CHAPITRE I. INTRODUCTION GENERALE : CONTEXTE ET PROBLEMATIQUE

I.1. Contexte.....	13
I.2. Problématique.....	15
I.3. Plan de la thèse.....	18

I.1. CONTEXTE

Cette thèse, au croisement de l'informatique, de la physique et de la biologie, a pour objectif d'élaborer un modèle de systèmes biologiques exploitable dans des études de radiobiologie menées *in silico*^{*}, en couplant son utilisation à des logiciels de simulation développés par la communauté des physiciens. La radiobiologie est la discipline qui fait le pont entre la physique et la biologie, elle qui a pour objectif de comprendre l'impact des rayonnements ionisants sur le matériel biologique.

Le vivant est soumis à l'irradiation par la radioactivité naturelle de l'environnement à laquelle nous sommes exposés en permanence. Si généralement cette radioactivité est faible, et que les systèmes biologiques savent s'en accommoder, elle peut être beaucoup plus importante et s'accompagner de dommages potentiellement très graves, que l'exposition soit volontaire, comme dans le cas d'une radiothérapie par exemple, ou involontaire en cas d'accident nucléaire, comme celui qui a touché la centrale Fukushima au Japon en mars 2011. De fait, la radioactivité représente un enjeu de santé publique par l'importance des risques générés par l'utilisation civile et militaire de l'énergie nucléaire, et par l'utilisation croissante de la radioactivité dans le traitement de certaines pathologies (cancers).

Les biologistes ont, de leur côté, la connaissance des systèmes auxquels ils s'intéressent, avec seulement la possibilité d'observation avant et après exposition à un stress radiatif. Les physiciens disposent d'outils de modélisation permettant la simulation des interactions particules-particules, dont certaines applications orientées vers la médecine permettent d'évaluer la quantité d'énergie mise en œuvre dans l'exposition à la radiation. En parallèle, il existe des modèles informatiques qui sont dédiés à l'étude de certains systèmes ou processus biologiques.

Il manque la synergie entre toutes ces composantes pour proposer un modèle capable de prédire les dommages consécutifs à une irradiation sur le matériel biologique, en fonction de la quantité d'énergie déposée.

Avec l'émergence des réseaux « -omiques » (génomique^{*}, métabolomique^{*}, transcriptomique^{*}...) et le développement de tests permettant d'obtenir des informations précises concernant la biochimie d'un individu, il devient même envisageable d'élaborer un modèle qui intègre des informations biologiques propres à l'individu et de les faire interagir avec les modèles de physique des particules pour faire des tests *in silico*. Un tel modèle serait d'importance dans le milieu médical, dans le contexte du développement de la médecine personnalisée, il fournirait un moyen de mieux prédire l'effet d'une radiothérapie sur un patient donné, en donnant l'opportunité aux praticiens d'adapter le traitement à la physiologie et à la physionomie de leurs patients, pour limiter les dommages sur les tissus environnant la zone à traiter.

^{*} Voir glossaire, page 199

Le premier jalon pour arriver à un tel outil passe par la maîtrise de la modélisation de ces phénomènes dans les êtres vivants parmi les plus simples qui soient, les bactéries. « Ce qui est vrai pour *Escherichia coli* est vrai pour un éléphant » (Jacques Monod).

Pour ce faire, nous nous projetons dans un cadre expérimental basé sur la mise en parallèle d'une approche *in vitro** et d'une approche *in silico* (Figure I.1) :

- *In vitro*, des lignées de cellules bactériennes sont cultivées et soumises à irradiation. On observe alors leur comportement. Le niveau le plus simple d'observation consiste en la remise en culture et en la comparaison des caractéristiques de la croissance avant et après irradiation ;
- *In silico*, on modélise des cellules bactériennes et un environnement dans lequel on les place en culture puis on simule la croissance des lignées cellulaires. On les soumet alors à un modèle d'irradiation et on simule le nouveau comportement du modèle cellulaire.

Les données obtenues *in vitro* servent d'étalon au comportement du modèle, en plus de permettre son paramétrage. Pour la modélisation de l'interaction entre les rayonnements ionisants et la matière, nous nous appuyons sur la communauté des physiciens et leur outil GEANT4, et notamment ses extensions dédiées à l'interaction entre les radiations et la molécule d'ADN.

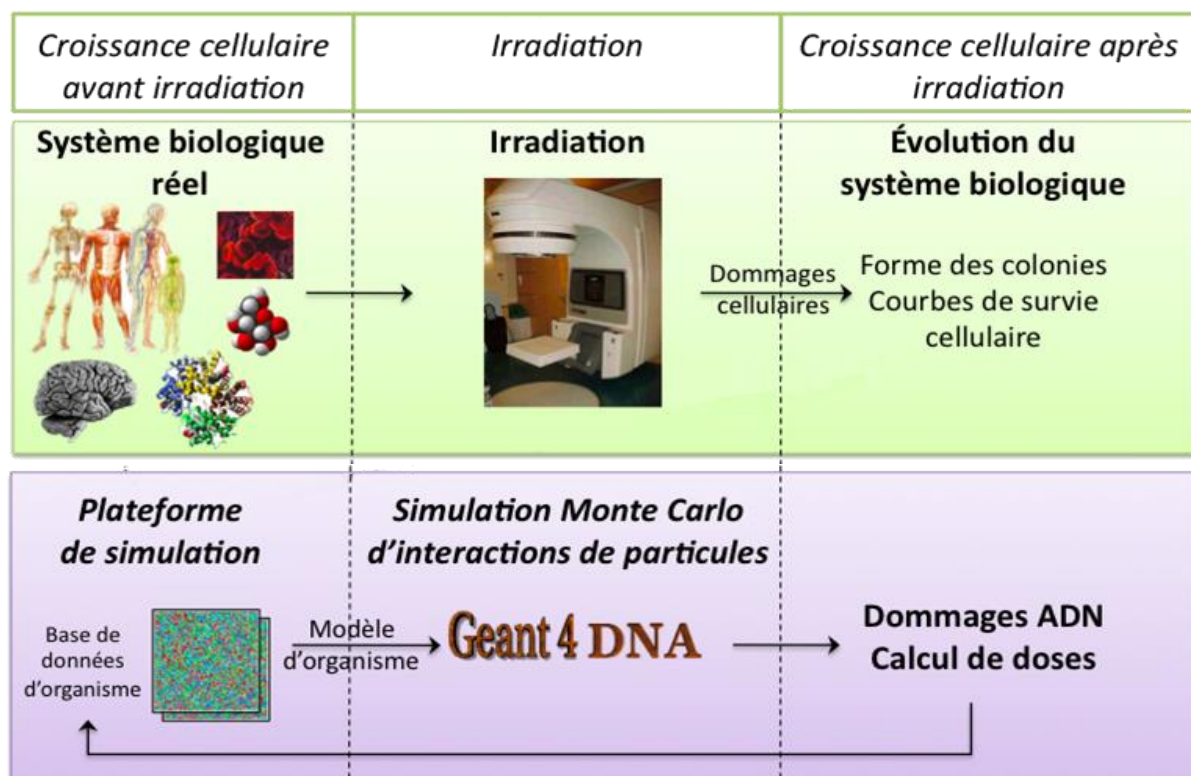


Figure I.1 : Mise en parallèle du cadre expérimental *in vitro*/*in vivo* et du cadre expérimental *in silico*. Le point d'articulation est une plateforme de modélisation et de simulation des systèmes biologiques.

* Voir glossaire, page 199

I.2. PROBLEMATIQUE

L'idée est donc de créer la synergie entre biologie et physique, avec l'informatique comme pont entre ces deux disciplines. C'est là le cœur des travaux présentés ici, fournir un cadre de modélisation et un modèle d'organisme qui puissent être mis en relation avec GEANT4 pour simuler l'irradiation, et dont le comportement soit étalonné et validé par les données obtenues *in vitro* ou *in vivo**.

Avant de proposer et de concevoir un cadre de modélisation, un état de l'art doit être réalisé. Cela nécessite auparavant de mieux définir le système biologique et la première étape est donc de préciser la nature du vivant, sa structure et sa composition.

La description des systèmes biologiques fait apparaître leur nature multi-échelle. Ces systèmes ont des propriétés qui s'observent à plusieurs échelles spatiales, de l'atome à plusieurs mètres, et temporelles, d'un millième de seconde à plusieurs millions d'années. Il y a une corrélation entre les échelles de temps et d'espace, les processus les plus rapides étant ceux qui se déroulent aux dimensions les plus réduites du système. Néanmoins, une échelle semble particulièrement importante, celle de la cellule qui est l'unité fonctionnelle du vivant. Ceci valide notre idée de commencer par nous intéresser aux bactéries dans notre projet de modélisation. En effet les bactéries sont des cellules, et plus précisément, elles sont des microorganismes unicellulaires, des organismes à part entière donc, mais qui ne sont constitués que d'une seule cellule. De plus les bactéries sont des cellules dites procaryotes, dépourvues de noyau et leur structure est plus simple que celle des cellules eucaryotes.

Une fois identifié cet élément central du fonctionnement du vivant, nous avons approfondi notre exploration jusqu'au niveau moléculaire. Ce niveau décrit tous les processus à l'œuvre dans une cellule pour lui permettre d'assurer sa survie. Ce niveau fait en plus le lien avec la physique, puisque c'est sur les molécules que vont s'exercer les effets de la radioactivité. À cette échelle, notre attention a particulièrement été attirée par les notions de génétique et de son support, l'ADN*. Nous avons alors considéré comment la molécule d'ADN code l'information génétique, comment elle est exprimée, et comment elle est transmise des ascendants aux descendants.

En observant la façon dont les cellules assurent leur survie, nous avons été amenés à considérer les milieux de vie des cellules. Nous avons identifié leurs propriétés pour pouvoir par la suite caractériser l'environnement modélisé pour notre modèle de cellule.

Dans le même temps, voir le fonctionnement des cellules aura mis en lumière les processus d'adaptation et d'évolution qui permettent à un organisme de s'accommoder de ces conditions environnementales. Ces processus qui prennent place sur des échelles de temps très brèves et très longues respectivement auront droit à notre considération, puisque nous souhaitons justement étudier l'effet du stress radiatif sur le vivant. Comprendre comment ils fonctionnent dans le système biologique nous permettra d'envisager leur modélisation au sein de notre cadre de travail.

* Voir glossaire, page 199

Cette description des systèmes biologiques nous amène à préciser quelques notions de radiobiologie et de culture cellulaire, puisque ces dernières définissent le contexte des expériences menées *in vitro*, et par conséquent, les éléments qu'il nous faut prendre en compte pour mener le parallèle *in silico*.

Dès lors que cette description nous a paru suffisante pour nous attaquer à la mise en place de notre environnement de modélisation et à la mise en œuvre de notre modèle de cellule, nous avons étudié les modèles existants.

Il existe deux grandes classes de modèles pour étudier les populations de cellules, les modèles à population et les modèles individus centrés. Les modèles à population ne rendent pas compte des spécificités individuelles des organismes, en traitant du comportement moyen qui s'observe dans une population. Notre objectif de modélisation dans une perspective radiobiologique ne peut se satisfaire de cette approche. En effet, il nous apparaît important de pouvoir suivre individuellement chaque cellule, l'intensité du stress qu'elle subit et comment elle y répond. Cette étude des caractéristiques de chaque individu est adressée par les modèles individu-centrés vers lesquels nous nous sommes orientés.

La variété de modèles individu-centrés pour les systèmes biologiques nous a conduits à étudier d'abord leur structure avant de décrire leur fonctionnement. Parmi les structures de modèles individu-centrés nous avons identifié, les automates cellulaires ont retenu notre attention pour leur simplicité de mise en œuvre et la richesse de leur fonctionnement.

Le fonctionnement des modèles s'est révélé plus difficile à explorer. Il existe une classification des modèles selon qu'ils exploitent ou non trois processus qualifiés d'élémentaires, la mort, la migration et la différenciation.

Au cours de la description des systèmes biologiques, nous avons identifié trois échelles d'espace, et autant de temps. D'un point de vue spatial, il nous apparaît nécessaire de pouvoir rendre compte de l'échelle moléculaire, où se déroulent les processus biochimiques et les effets des radiations sur le vivant, de l'échelle cellulaire, à cause du rôle clé des cellules dans le fonctionnement du vivant, et de l'échelle de la population, où s'observent les interactions entre les cellules. D'un point de vue temporel, les échelles qui nous apparaissent indispensables concernent les échelles de temps très courts permettant d'observer les processus biochimiques, l'échelle de la vie de la cellule, pour observer son adaptation à son environnement, et le temps qui couvre plusieurs générations, pour étudier les processus évolutifs. Cette classification basée sur migration, différenciation et mort nous apparaît dès lors trop réductrice ne permettant de rendre compte que de temps correspondant à l'adaptation.

Nous proposons de classer les modèles selon processus biochimiques, adaptation et évolution qui correspondent respectivement à l'échelle subcellulaire, cellulaire et d'une population cellulaire. A la lumière de cette nouvelle classification nous avons étudié des modèles correspondants. Il ressort de cette étude que chaque modèle n'est prévu que pour simuler des processus qui sont observés à l'échelle qui leur correspond, et nous n'avons pas identifié de modèle qui permette d'englober les trois catégories, ce qui correspond pourtant au besoin de notre cadre radiobiologique.

En parallèle à l'exploration des modèles existants, nous avons exploré les environnements logiciels, qui permettent de fournir le cadre de simulation à ces modèles. Une différence importante nous est apparue dans la conception de ces environnements logiciels, les plateformes. D'un côté il y a les plateformes qui sont faites pour un modèle, elles en sont indissociables, nous les qualifions de dédiées. En opposition, il y a les plateformes que nous qualifions de génériques, qui sont conçues autour d'un formalisme et qui permettent de faire jouer différents modèles écrits selon ce formalisme.

Là encore, nous n'avons pas trouvé d'environnement logiciel répondant à nos attentes. Les plateformes dédiées n'offrent pas la souplesse qui serait requise pour développer un modèle différent de celui pour lequel elles sont conçues. Les plateformes génériques, même si elles permettent de faire jouer plusieurs modèles, sont conçues autour de formalismes eux-mêmes écrits pour traiter des processus spécifiques comme la morphogénèse. De fait, les plateformes génériques ne se prêtent pas à une simulation générale de croissance cellulaire dans un milieu stressant.

À l'issue de cet état de l'art, nous avons donc été amenés à concevoir un modèle et une plateforme pour l'accueillir. Le développement de cette dernière est contraint par une volonté de rester sur un environnement général, qui pourra évoluer à mesure que le modèle s'affinera, et qui permettra de faire jouer plusieurs modèles pour répondre aux conditions expérimentales. Nous avons réduit notre champ de développement aux modèles individuentrés, et particulièrement aux automates cellulaires comme structure de modèles, aussi le développement de la plateforme s'est-il orienté vers ce type de modèles. L'architecture générale proposée gère une population constituée d'individus placés dans un environnement. Pour valider le comportement de la plateforme et vérifier qu'elle n'induit pas de biais, nous avons, dans un premier temps, vérifié sa capacité à faire jouer correctement des automates cellulaires de la littérature (Le Jeu de la Vie et *Wireworld*), et un modèle de la littérature, *Daisyworld*, que nous avons spécifié sous forme d'un automate cellulaire.

Après validation du cadre logiciel, nous avons développé un modèle de cellule pensé pour prendre en considération les phénomènes qui se passent à l'échelle moléculaire, avec les informations génétiques, d'une part, et le déroulement des processus biochimiques d'autre part. Les processus biochimiques dépendent des caractéristiques génétiques, et ces dernières peuvent être modifiées par les processus biochimiques.

L'observation du comportement du modèle se fait à l'échelle de la population et il résulte des comportements individuels des cellules. Les cellules sont capables d'interagir et de percevoir de l'information de l'environnement, ce qui nous permettra de modéliser les conditions environnementales voulues, qu'elles présentent un stress ou non. Par les informations génétiques qu'elles portent et par les processus biochimiques qui leur correspondent, les cellules pourront exploiter cet environnement et s'adapter, en apportant une réponse à un stress. Enfin, suivre le comportement de cette population dans le temps peut conduire à l'observation de phénomènes évolutifs, puisque le patrimoine génétique n'est pas figé, une mutation peut survenir, et le stress environnemental peut conduire à la sélection d'une lignée.

Avant d'envisager de modéliser les cellules dans un environnement stressant, il nous faut maîtriser la modélisation de leur comportement dans un milieu « normal ». Avec le modèle en main, nous avons cherché des informations biochimiques quantifiées dans la littérature pour calibrer et valider la bonne reproduction *in silico* d'expériences menées *in vitro*, à l'aide des valeurs de paramètres issues de ces dernières.

I.3. PLAN DE LA THESE

Nous concluons cette partie en détaillant le plan du mémoire.

Après ce premier chapitre qui nous a permis de décrire le contexte de ces travaux et de définir la problématique qui nous intéresse, le second chapitre sera consacré à l'étude des systèmes biologiques, à la description de leurs propriétés et de leurs fonctions. Notre description des systèmes biologiques commencera avec des généralités sur le fonctionnement du vivant. Cela nous conduira à voir la complexité des systèmes biologiques, et à identifier les niveaux d'observation de ces systèmes, et la façon dont ils sont interconnectés. Dans le même temps, la distinction entre système biologique et système vivant nous conduira à une réflexion sur la définition de la vie. Cette dernière fait entre autres intervenir la notion de reproduction et d'hérédité. Après description de cette dernière, nous nous intéresserons à son fonctionnement et à son support, le matériel génétique. La transmission d'informations génétiques se fait lors de la multiplication des cellules qui se retrouvent être au cœur du fonctionnement du vivant dans les échelles de temps et d'espace. La cellule sera alors l'objet de notre attention, et nous nous attacherons à décrire sa structure, avec les différents types de cellules qui existent, puis nous considérerons son fonctionnement. Nous parlerons alors du cycle de vie des cellules, puis de la manière des cellules d'assurer leur subsistance. Ceci nous amènera à considérer ce qui constitue un environnement viable pour une cellule et nous identifierons les caractéristiques de leur milieu de vie. De ces caractéristiques, nous verrons que certaines peuvent être source d'un stress sur le vivant. Celui-ci peut alors y apporter une réponse adéquate, c'est l'adaptation, et à travers les capacités d'adaptation s'exerce la sélection naturelle. Nous clôturerons ce second chapitre en précisant le cadre expérimental qui accompagne ces travaux.

Le troisième chapitre sera consacré à la présentation de l'état de l'art réalisé. Au cours de celui-ci nous serons amenés d'abord à comparer les deux grandes catégories de modèles : les modèles à population et les modèles individu-centrés. Nous présenterons quelques modèles à population puis nous focaliserons notre attention sur les modèles individu-centrés. Après quelques exemples de ces derniers, nous présenterons les automates cellulaires, les modèles de Potts, les systèmes Lattice-Gas et les systèmes multi-agents. C'est sur les automates cellulaires que s'est arrêté notre choix et nous justifierons ce choix par la richesse de leur fonctionnement et leur simplicité de mise en œuvre. Nous passerons alors en revue des modèles individu-centrés et ce sont les automates cellulaires employés pour modéliser le vivant. Pour explorer cette partie de l'état de l'art, nous proposons une classification de ces modèles selon les échelles de temps qu'ils prennent en compte ce qui nous permet de passer outre les limitations de la classification existante basée sur les processus de mort, migration et différenciation. Nous étudierons quelques-uns des modèles

adressés aux processus biochimiques, à l'adaptation puis à l'évolution. Suite à tout ceci, nous aborderons l'aspect logiciel qui accompagne ces modèles et nous parlerons des plateformes, en distinguant celles que nous qualifions de dédiées, dépendantes du modèle pour lequel elles sont conçues, des plateformes que nous qualifions de génériques, architecturées autour de formalismes. À l'issue de cet état de l'art, nous concluons à l'absence d'un modèle et d'une plateforme répondant à notre problématique et pouvant s'inscrire dans la durée dans le cadre du contexte de recherche que nous avons décrit.

Le quatrième chapitre traitera de la solution que nous proposons. Dans un premier temps, nous aborderons l'aspect logiciel et nous expliquerons la plateforme que nous avons développée. Nous présenterons d'abord les concepts qui ont guidé son élaboration ainsi que son architecture, puis nous détaillerons les étapes de validation de son comportement. Pour ce faire, nous présenterons la mise en œuvre d'automates cellulaires de la littérature, le Jeu de la Vie et Wireworld, et nous vérifierons que les résultats que nous obtenons sont en accord avec leur description. Ensuite, nous nous proposerons la mise en œuvre sous forme d'un automate cellulaire d'un modèle de la littérature, Daisyworld. Nous le traiterons de deux façons différentes, d'abord en négligeant l'élément perturbateur du système modélisé, puis en le prenant en compte, et nous vérifierons également que les résultats sont conformes à ceux connus pour ce modèle. Ce sera alors pour nous l'occasion de faire un parallèle avec d'autres travaux menés au sein de l'équipe pour l'identification de règles de fonctionnement d'automates cellulaires. Une fois les propriétés de la plateforme validées, nous ferons la proposition d'un modèle de cellule. Nous détaillerons les hypothèses de modélisation qui ont guidé son élaboration, puis nous ferons état de ses fonctions. Nous ferons également un point pour aborder les paramètres qui servent à étalonner son fonctionnement.

Le cinquième chapitre sera consacré à la vérification du fonctionnement du modèle, à sa robustesse et à son efficacité. Au cours de ces étapes, nous utilisons un jeu de paramètres arbitraire. Nous décrivons les limites et contraintes que nous avons rencontrées et les ajustements qui ont été nécessaires. Une fois certains que les résultats que nous obtenions étaient bien ceux que nous devions obtenir, nous avons procédé à l'application de paramètres significatifs en biologie sur les fonctions du modèle. Nous présenterons les résultats obtenus, en comparant les résultats obtenus *in silico* aux résultats obtenus *in vitro* dans la littérature. Notre modèle démontrera alors qu'il permet de prendre en considération des données biochimiques précises dans son fonctionnement.

Nous rappellerons les résultats présentés pour conclure ce document. Nous nous projetterons dans le contexte expérimental et décrivons comment les nouvelles données ou fonctions pourront être intégrées au modèle, ou prises en compte par celui-ci pour modéliser la culture de cellules en culture, dans un environnement sain ou stressant, ou dans le cas d'un stress temporaire. Nous évoquerons également les évolutions à apporter pour permettre d'affiner le comportement du modèle et pour permettre la communication avec les outils comme GEANT4. À l'issue de ces travaux nous aurons posé les bases d'un système qui permettra à terme de disposer *in silico* d'un modèle qui peut être paramétré et étalonné à partir des données d'expériences *in vitro*.

CHAPITRE II. PROPRIETES ET MECANISMES DES SYSTEMES BIOLOGIQUES

II.2. Généralités.....	24
II.3. L'hérédité et la transmission de patrimoine génétique.....	29
II.3.1. Principes de l'hérédité.....	29
II.3.2. Notions de génomique.....	32
II.3.2.1. Structure du génome.....	32
II.3.2.2. Expression du génome.....	34
II.3.2.3. Transmission du génome.....	36
II.4. L'unité fonctionnelle du vivant : la cellule.....	38
II.4.1. Structure des cellules.....	38
II.4.1.1. Aspects généraux.....	38
II.4.1.2. Présentation de la cellule procaryote.....	40
II.4.1.3. Présentation de la cellule eucaryote.....	41
II.4.2. Vue d'ensemble de la vie des cellules.....	43
II.4.2.1. Croissance.....	43
II.4.2.2. Division.....	44
II.4.2.3. Mort.....	45
II.4.3. Vue d'ensemble du fonctionnement cellulaire.....	45
II.4.3.1. Métabolisme.....	45
II.4.3.1.1. Catabolisme.....	47
II.4.3.1.2. Anabolisme.....	47
II.4.3.2. Échanges entre la cellule et son environnement, notions de transport.....	48
II.4.3.2.1. Transport passif.....	49
II.4.3.2.2. Transport actif.....	49
II.5. Environnement.....	50
II.6. Notions d'adaptation et d'évolution.....	51
II.6.1. Adaptation.....	52
II.6.2. Evolution.....	52
II.7. Cadre expérimental.....	54
II.7.1. Environnements expérimentaux, notion de culture cellulaire.....	55
II.7.1.1. Culture en milieu solide.....	56
II.7.1.2. Culture en milieu liquide.....	57
A. Phase de latence.....	58
B. Phase exponentielle.....	58
C. Phase stationnaire.....	59
D. Phase de déclin.....	59
II.7.2. Radiobiologie.....	59
II.8. Conclusion.....	62

II.1. INTRODUCTION

L'objectif de ce chapitre est de décrire le système auquel nous allons nous intéresser, le système biologique, ses caractéristiques et ses fonctions, puis nous aborderons le contexte spécifique qui nous intéresse, la radiobiologie.

Pour cela nous commencerons par décrire le vivant dans son ensemble pour arriver à son entité unitaire de fonctionnement, la cellule. En passant, nous ferons la distinction entre les différents types d'organismes et entre système biologique et système vivant.

Le vivant est observé depuis toujours à l'échelle qui nous est accessible directement, les organismes qui nous entourent, leurs interactions, leurs comportements. Plus récemment, les progrès technologiques nous ont permis d'explorer la nature plus en détail. Découvrant son unité fonctionnelle, la cellule, puis en observant de plus en plus finement, ce sont les processus chimiques qui sous-tendent la vie qui se sont fait jour. La notion d'hérédité, qui était connue de longue date à travers ce qui s'observe au niveau des organismes ou des populations a alors révélé ses secrets, et le voile s'est levé sur son fonctionnement. Notre exploration des systèmes biologiques suivra cette chronologie.

Nous parlerons d'abord du vivant dans son ensemble et réduirons notre champ d'observation à la cellule et à ce qui se passe en son sein. Nous ferons ensuite une parenthèse pour parler d'hérédité, ce qui nous permettra d'aborder les notions de génétique à l'échelle d'un individu et d'une population. Ces notions nous amèneront à nous intéresser au génome*. Nous décrirons sa structure, son expression et sa transmission.

Il nous faudra alors considérer le fonctionnement des cellules. Nous commencerons par l'observer dans sa globalité, en détaillant la vie des cellules, puis nous explorerons un peu plus précisément la façon dont une cellule échange de la matière avec son environnement, puis c'est le métabolisme qui sera à son tour détaillé, en abordant ses deux composantes que sont le catabolisme et l'anabolisme. En parlant du génome nous aurons été amenés à considérer l'existence de différents types de cellules que nous allons caractériser succinctement.

Étudier le métabolisme et la croissance cellulaire nous fera apercevoir les exigences des cellules et nous nous attacherons alors à comprendre l'environnement d'une cellule et la relation qu'elle entretient avec lui. Nous verrons alors les caractéristiques du milieu naturel et ce qu'il faut en comprendre comme informations pour être en mesure de mettre des cellules en culture.

La génétique et les contraintes présentes dans l'environnement seront à la base de l'adaptation, cette dernière, par le biais de l'hérédité sera à l'origine de l'évolution. Adaptation et évolution seront alors le sujet de notre attention et nous les présenterons telles qu'elles ont été observées à l'échelle des organismes et des populations, avant de les observer également sous l'angle de la génétique.

* Voir glossaire, page 199

Pour finir, nous aborderons deux notions particulières ; la première concernera la caractérisation de la croissance cellulaire et comment elle peut être étudiée, afin de mieux cerner le cadre expérimental de ces travaux ; la seconde, la radiobiologie est une discipline, au croisement de la biologie et de la physique, qui nous amènera à nous interroger sur les effets de la radioactivité sur le vivant.

II.2. GENERALITES

La biologie couvre une très large palette de disciplines qui s'intéressent à des domaines de dimensions et de niveaux d'organisation très différents, de la biosphère, qui représente le niveau d'organisation le plus complexe, aux plus petites molécules produites ou utilisées par des organismes vivants, en passant par l'étude des organismes, de leur structure, de leur fonctionnement et de leur développement.

Tous ces domaines sont imbriqués les uns dans les autres comme des poupées gigognes. Ainsi, la biosphère considère comme un tout l'ensemble des écosystèmes, ces écosystèmes représentent l'ensemble des organismes qui les habitent, les relations et échanges qui s'établissent entre eux (chaînes alimentaires, symbiose*, commensalisme*...), et avec leur environnement (niche écologique). « ...tout être organique est relié [...] à tous les autres êtres organiques avec lesquels il entre en compétition pour la nourriture ou l'habitat, ou auxquels il doit échapper, ou dont il fait ses proies » (Darwin).

Classiquement, on considère deux grands groupes d'organismes, ceux dits inférieurs et ceux dits supérieurs. Les organismes inférieurs sont représentés essentiellement par les procaryotes* et les protistes*. Les procaryotes regroupent les eubactéries* (ou, communément, les bactéries) et archaeobactéries*. Nous ne parlerons pas davantage de ces dernières qui sont un peu à part dans le sens où elles occupent souvent des niches écologiques présentant des caractéristiques extrêmes.

Les bactéries sont souvent le dernier maillon de la chaîne alimentaire, en étant les derniers décomposeurs de la matière organique. Certaines bactéries sont également autotrophes, capables de transformer de la matière minérale en matière organique grâce à l'énergie lumineuse ou l'énergie chimique. De fait, certaines bactéries peuvent également constituer le premier maillon de la chaîne alimentaire.

Par opposition aux procaryotes, cellules dépourvues de noyau, les autres cellules sont dites eucaryotes*, dotées d'un noyau.

Les protistes sont des organismes unicellulaires eucaryotes qui se nourrissent de débris organiques, de bactéries ou d'algues microscopiques. Ils ont un rôle de décomposeur et représentent souvent des maillons de chaînes alimentaires. Les organismes dits supérieurs seront représentés par les végétaux, les champignons et les animaux. Ils sont tous constitués de cellules eucaryotes. Les végétaux sont des organismes autotrophes qui réalisent la photosynthèse*, ils puisent du gaz carbonique dans l'atmosphère, des sels minéraux et de l'eau dans le sol, et ils transforment ces substances minérales qu'ils captent dans leur

* Voir glossaire, page 199

environnement en matière organique, des sucres, grâce à l'énergie lumineuse. Ces sucres seront ensuite utilisés pour élaborer les autres constituants que sont les protéines, les lipides,... Puisqu'ils produisent la matière organique à partir de la matière minérale, ils sont souvent à la base des chaînes alimentaires.

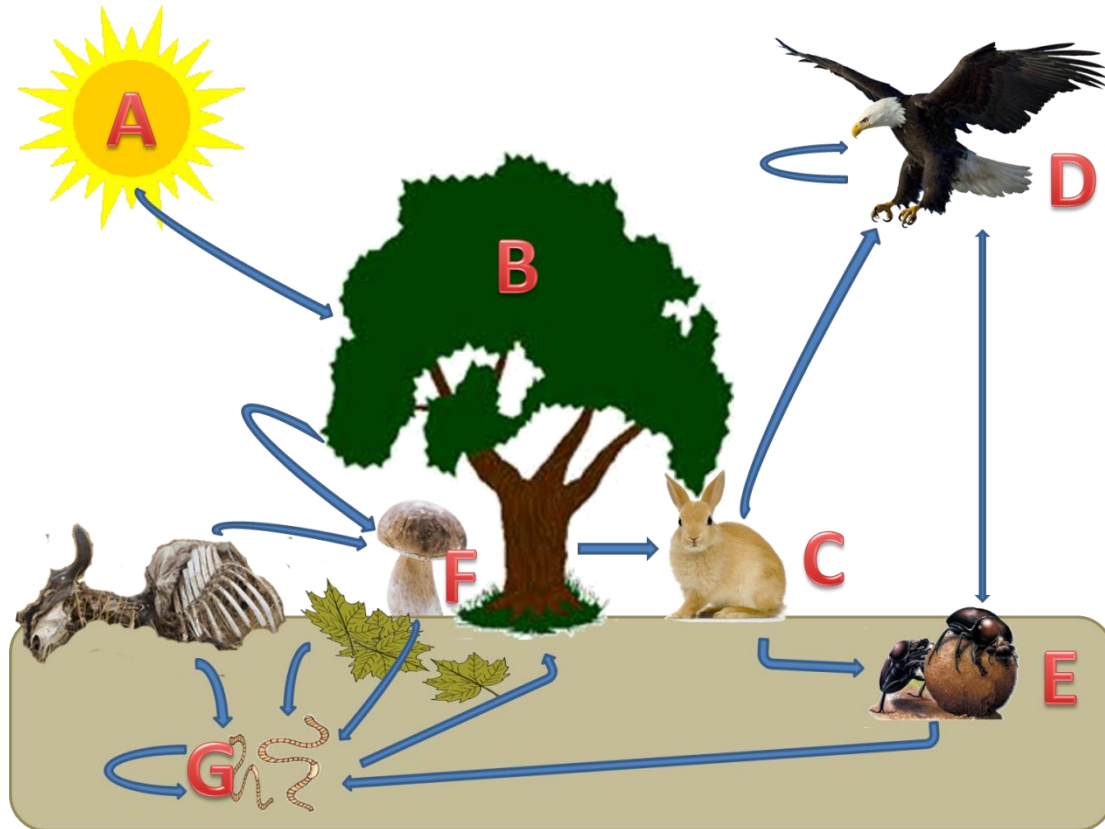


Figure II.1 : Schéma simplifié d'un réseau trophique, A le soleil, B les végétaux, C les herbivores, D les prédateurs, E les détritivores, F les champignons, G les décomposeurs.

Le soleil (A) fournit l'énergie aux végétaux et autres organismes photosynthétiques (B). Ce sont des producteurs primaires. Ces derniers sont mangés par des animaux herbivores (C) qui sont des producteurs secondaires. Les animaux herbivores sont mangés par des prédateurs (D) qui sont également des producteurs secondaires. Les prédateurs peuvent avoir pour proie des prédateurs, ils sont alors qualifiés de superprédateurs. Les déjections des herbivores et des prédateurs servent de nourriture à des détritophages (E) qui sont des décomposeurs primaires. Les matières animales ou végétales, mortes ou vivantes, servent de support au développement des champignons (F). Les animaux morts, la matière végétale morte, les champignons morts sont décomposés par les décomposeurs terminaux (G), parmi lesquels on va trouver des animaux, des bactéries, des champignons, etc... Les décomposeurs terminaux rejettent des substances minérales, qui servent de nutriments aux végétaux (B)

Les champignons sont des organismes qui se nourrissent de matière organique, ils sont dits hétérotrophes. Ils ont souvent un rôle de décomposeurs, permettant de recycler de la matière organique, comme des végétaux morts par exemple, en la dégradant en gaz carbonique, qui retourne dans l'atmosphère, et en minéraux, qui restent dans le sol. Ce rôle les place en fin de chaîne alimentaire. Ils peuvent aussi faire partie des maillons d'une chaîne alimentaire s'ils sont mangés par des animaux. Les animaux sont des organismes hétérotrophes. Ils se nourrissent de matière organique qui peut être d'origine végétale, pour

les herbivores, ou animale, pour les carnivores, avec toutes les variantes que l'on peut envisager : xylophages qui mangent le bois, charognards qui se nourrissent de chair morte, coprophages qui se nourrissent d'excréments, etc... Les animaux représentent la plupart des maillons intermédiaires des chaînes alimentaires. Les animaux ont également un rôle dans le recyclage de la matière organique, par leurs déjections, et certains ont un rôle de décomposeurs.

Les chaînes alimentaires sont aussi appelées chaînes trophiques. Dans un écosystème, les différentes chaînes trophiques sont interconnectées pour former un réseau trophique (Figure II.1). Dans ce réseau, les déchets des uns sont les nutriments des autres, ce qui permet au cycle de la vie de se perpétuer en un cycle vertueux. « La nature entière est une conjugaison du verbe manger, à la forme active et à la forme passive » (Inge).

Pour chaque type d'organisme on peut s'intéresser à sa structure, à son fonctionnement, à son organisation et à son développement. Historiquement, seuls des organismes pluricellulaires, de grande taille, étaient décrits. On connaissait des microorganismes pour leurs effets, mais pas par leur observation directe, ou alors à travers certaines structures de taille suffisamment importante pour être observées à l'œil nu. C'étaient par exemple les biofilms, pellicules visqueuses qui se développent à la surface d'une eau croupissante, ou les tâches colorées qui accompagnent la pourriture. Souvent on les appelait miasmes.

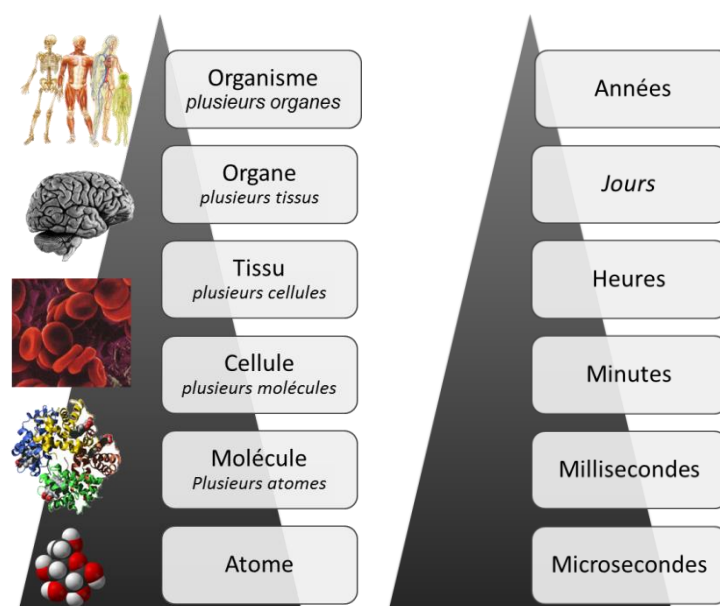


Figure II.2 : Les niveaux d'observation d'un système biologique, par exemple un être humain, et les échelles de temps auxquelles ils peuvent correspondre.

L'observation des organismes supérieurs a permis de constater qu'ils sont dotés de plusieurs organes ayant chacun une fonction spécialisée. Ces organes sont communs au sein d'un règne, ainsi, les animaux étaient connus pour se déplacer à l'aide de pattes, parmi eux on en compte dotés d'un exosquelette, les insectes, alors que d'autres ont un squelette interne, les vertébrés. Parmi ces derniers, certains pondent des œufs, dont sortent les petits, alors que d'autres donnent directement naissance à leurs petits, etc... Plus tard, on s'est aperçu que chaque organe est lui-même structuré de plusieurs tissus. Et avec l'avènement de la

microscopie, on a découvert que chaque tissu est composé de cellules spécialisées travaillant ensemble. Avec les techniques plus récentes de la chimie et la sophistication des techniques d'imagerie, l'exploration des échelles du vivant s'est poursuivie et l'observation a montré qu'une cellule regroupe une grande quantité de molécules qui sont des édifices poly-atomiques.

La structure d'un organisme peut ainsi être observée à plusieurs échelles, comme montré sur la Figure II.2. On note un parallèle entre les échelles spatiales observables, et les échelles de temps qui leur sont reliées. Les phénomènes à l'échelle atomique se déroulent sur des temps de l'ordre de la microseconde voire moins, alors qu'à l'autre extrémité, la vie de l'organisme s'observe sur des durées allant de quelques jours à plusieurs années. Quelle que soit l'échelle considérée, un bon fonctionnement est lié à la bonne marche des parties constitutives. Par exemple, si l'on considère l'être humain, les systèmes squelettique, musculaire, nerveux, circulatoire, etc... doivent fonctionner de concert. Chacun de ces systèmes est constitué de plusieurs organes et il suffit de la défaillance d'un seul pour faire défaillir le système et, par réaction en chaîne, l'ensemble de l'organisme (Figure II.3).

L'organisation et le développement d'un organisme sont liés. Le développement va s'intéresser au devenir de l'organisme, de la naissance à la senescence et l'organisation sera une conséquence du développement. Nous autres êtres humains sommes caractérisés par un axe de symétrie qui sépare notre droite de notre gauche. Nos organes sont organisés selon cet axe, sur un axe dorso-ventral et sur un axe antéro-postérieur. Ces axes se mettent en place dès le développement embryonnaire et vont servir de référentiel au bon positionnement des différents systèmes. Ainsi le tractus digestif commence au niveau de la tête par la bouche, prolongée par l'œsophage, pour arriver à l'estomac, puis viennent l'intestin grêle et le gros intestin, avant d'arriver au rectum. Un mauvais positionnement se traduira par une malformation, et la plupart des malformations empêcheront le développement correct de l'organisme, menant même souvent à sa mort prématurée.

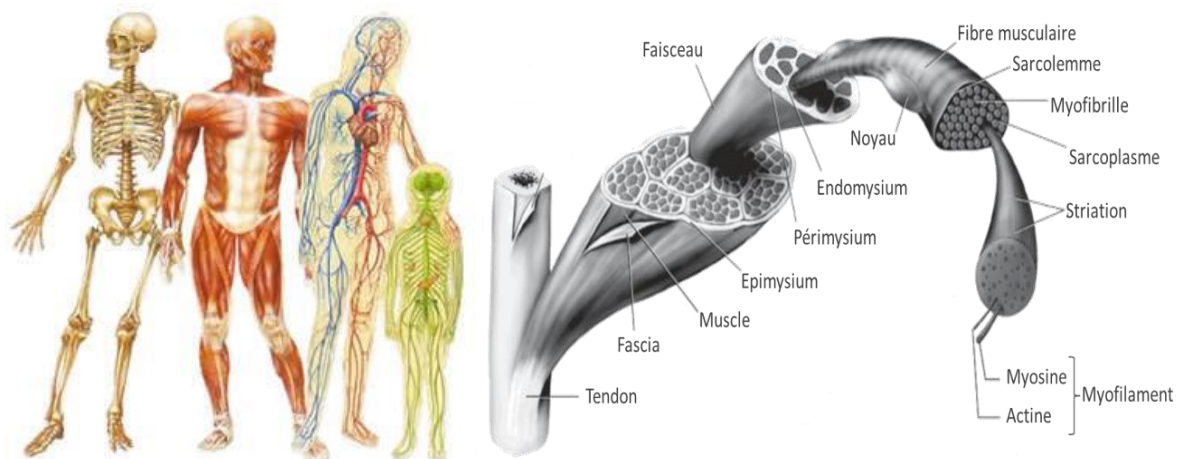


Figure II.3 : A gauche : Le système biologique est composé de plusieurs éléments en interaction, Le corps humain de la tête aux pieds (éditions atlas) ; à droite : un système fait intervenir plusieurs niveaux d'organisation, par exemple la structure du muscle (W. Whiting et S. Rugg, 2005)).

Aux échelles subcellulaires, les molécules sont plus ou moins complexes et sont des assemblages d'atomes plus ou moins nombreux. La multitude de molécules qui composent

une cellule obéit à la même règle que décrite précédemment pour les organes et organismes, une molécule mal formée ou abîmée nuit au bon fonctionnement de la cellule, et peut même mener à l'arrêt de son fonctionnement et sa mort.

Si l'on peut parler de système biologique lorsque l'on parle des molécules impliquées dans le vivant, on ne peut parler de système vivant lorsque l'on traite des échelles moléculaires ou atomiques. En effet, la définition couramment admise du vivant, comme par exemple celle trouvée dans un dictionnaire tel que le Larousse, implique la capacité à croître, à échanger avec l'environnement, pour se procurer de l'énergie ou pour renouveler ses constituants, et à se reproduire, tout en étant capable de résister à des changements survenus dans son environnement. Par extension, le vivant définit l'ensemble des processus entre la conception et la mort d'un organisme.

Cette définition régulièrement remise en question est souvent enrichie. Elle permet d'opposer vivant à inerte. Elle fait souvent débat car se posent des problèmes comme celui des virus. Les virus sont des particules de quelques nanomètres à quelques dizaines de nanomètres qui contiennent un génome et des enzymes, mais sont incapables d'avoir une vie autonome au sens de la définition énoncée précédemment. Ils doivent contaminer une cellule et détourner sa machinerie à leur profit pour proliférer.

Longtemps, la frontière du vivant prenait en compte une limite de taille, les cellules les plus petites connues mesurant 100 nm de diamètre. Mais cette distinction est devenue obsolète avec les découvertes récentes qui ont mis à jour certains virus, qualifiés de géants, dont la taille est de plusieurs centaines de nanomètres, soit plus que beaucoup de cellules, et pratiquement la taille courante des bactéries.

La distinction entre vivant et inerte est loin d'être parfaitement tranchée. De par l'histoire, Aristote a été le premier à proposer cette distinction en considérant le vivant comme un objet en mouvement. Il distinguait trois types de mouvement :

- Le mouvement qualitatif, qui concerne la transformation des organismes ;
- Le mouvement quantitatif, qui concerne la croissance des organismes ;
- Le mouvement local, qui concerne le déplacement des organismes. Ce dernier est le propre des animaux.

Il aura fallu attendre le 19^{ème} siècle pour que la définition actuelle du vivant commence à se construire. La découverte des microorganismes et le problème posé par la théorie de la génération spontanée confrontaient les savants au besoin d'avoir une définition du vivant sur laquelle se baser pour classer la nature et établir son ordre en sortant du cadre théologique qui subordonnait le monde à la volonté d'une intelligence supérieure. Les travaux de Louis Pasteur ont joué à ce niveau un rôle déterminant.

Si l'on s'en tient à la définition actuelle, le monde vivant commence à l'échelle de la cellule et s'achève à l'échelle de l'organisme, voire à l'échelle de la population. En deçà de l'échelle cellulaire, les échelles moléculaires et atomiques seront ainsi considérées du domaine de l'inerte. Cependant, si ces échelles ne peuvent être considérées dans le système vivant, elles auront un grand intérêt dans les disciplines de la biologie, dès lors que les molécules étudiées sont importantes pour le vivant et ces échelles seront donc incluses dans les systèmes biologiques.

Le problème des microorganismes et de la génération spontanée transparaissait à travers l'idée d'hérédité. En effet, il était connu qu'un individu d'une espèce donnée avait une descendance de la même espèce, "les chiens ne font pas des chats" dit la sagesse populaire, mais comment justifier alors que là où il n'y avait visiblement pas d'organisme, un miasme puisse se développer. Pasteur affirmait qu'un être vivant ne peut venir que d'un être vivant et la découverte des bactéries lui donna raison. L'hérédité se trouva alors confirmée et se révéla une vérité pour tous les organismes, et avec les progrès de la biologie moléculaire on put comprendre son support.

II.3. L'HEREDITE ET LA TRANSMISSION DE PATRIMOINE GENETIQUE

II.3.1. PRINCIPES DE L'HEREDITE

Nous venons de voir que l'entité la plus simple du vivant est la cellule. Tous les êtres vivants sont constitués d'au moins une cellule. Un organisme constitué d'une cellule unique sera dit unicellulaire, par opposition aux organismes pluricellulaires.

Tous les êtres unicellulaires donnent une descendance unicellulaire, et de la même manière tous les êtres pluricellulaires ont une descendance constituée d'organismes pluricellulaires. Et trivialement, on ajoutera que les représentants d'une espèce donnée ne donnent naissance qu'à des descendants de la même espèce.

L'observation peut même s'affiner au regard de caractères plus précis. C'est ce qui permit à Gregor Mendel de concevoir les bases de la génétique. Ce moine qui se passionnait pour la botanique avait observé que féconder des plans de pois jaunes et ridés avec du pollen issu de plans similaires donnait des plans aux mêmes caractéristiques, alors que féconder ces mêmes plans avec du pollen issu de plans de pois verts et lisses donnait des plans de pois verts et lisses. Il poussa même l'observation en constatant qu'apparier ces derniers plans de pois donnait pour un quart des plans de pois jaunes et ridés et pour trois quart des plans de pois lisses et verts.

À travers ces observations, il posa les règles de génétique qui sont aujourd'hui connues sous le nom de génétique mendélienne et il énonça le principe d'hybridation. Ceci permit d'expliquer ce qui était considéré comme des bizarreries, tel par exemple le mulet, stérile, issu de l'accouplement d'un âne et un cheval.

Il manquait la compréhension du support de cette information génétique pour finir de comprendre le fonctionnement de l'hérédité. Il fallut plusieurs développements successifs pour en arriver là. La découverte des cellules et de leur mode de multiplication, la découverte de la chimie des cellules, l'identification de la molécule d'ADN et la caractérisation de la structure de l'ADN figurent parmi les plus importantes.

On sait aujourd'hui que l'hérédité se fait grâce à la transmission de patrimoine génétique. Chez les organismes unicellulaires qui se reproduisent en se divisant, il y a d'abord copie du génome pour que chaque cellule fille puisse hériter d'une version, on parle de reproduction

asexuée. Par opposition, chez les organismes supérieurs qui recourent à la reproduction sexuée, il y a production de gamètes* qui ne portent que la moitié de l'information génétique, le nouvel individu est formé lors de la fécondation qui met en commun la moitié du patrimoine génétique obtenu du gamète mâle et la moitié du génome donné par le gamète femelle. Les gamètes sont aussi appelés cellules sexuelles.

Les expériences de Mendel avaient commencé à mettre en évidence une relation de dominance de certains caractères sur d'autres. Là encore, seules les technologies modernes permettent d'entrevoir toute la complexité des mécanismes qui régissent la dominance ou la récessivité d'un caractère.

Nous n'entrerons pas dans les détails mais nous précisons que la régulation de l'expression de l'un ou l'autre caractère se fait par des interactions chimiques qui ont lieu au sein de la cellule, ou par le biais de modifications chimiques sur l'ADN qui permettent de moduler l'expression de certains caractères. Tous les mécanismes en jeu sont encore loin d'avoir dévoilé toute leur subtilité, aussi pour la suite de ces travaux nous ne considérerons que les règles de base de la génétique, qui suffisent pour expliquer qu'un individu d'une espèce donnée donne naissance à un individu de même espèce, et que la modification subie par un individu au cours de sa vie ne se transmet pas à sa descendance.

On sait certains caractères dominants, c'est-à-dire qu'ils s'expriment en tous cas, d'autres sont récessifs, ils ne s'expriment que s'il existe chez un individu deux copies du caractère. Lorsque les deux copies sont légèrement différentes, le caractère dominant s'exprime, si les deux caractères sont dominants, ils peuvent s'exprimer dans le même temps. Il peut donc y avoir de subtiles nuances entre les ascendants et les descendants. On parle de génotype* pour décrire les caractères génétiques portés par un individu, et de phénotype* l'ensemble des caractères exprimés par un individu.

Par exemple, chez les humains, le gène "œil bleu" est un caractère récessif. Un individu ne peut avoir les yeux bleus que s'il porte deux copies du gène "œil bleu", par déduction, il hérite de ce caractère de chacun de ses parents. Ce qui signifie que chaque parent portait au moins une copie du gène "œil bleu". Chacun des parents portait au moins une copie du gène œil bleu, mais l'autre version du gène qu'ils possédaient ne peut être déduite. Chaque parent peut ainsi être porteur du gène "œil bleu" en deux exemplaires, et il a alors les yeux bleus, ou il peut porter le gène "œil foncé" et auquel cas, il a les yeux foncés. Ainsi, un couple aux yeux bleus ne peut avoir que des enfants aux yeux bleus, alors qu'un couple aux yeux foncés peut avoir des enfants aux yeux clairs ou aux yeux foncés selon les gènes dont eux-mêmes sont porteurs.

Lorsque les deux versions du gène sont les mêmes on parle d'homozygotie*. Un individu qui a le caractère "œil bleu" est homozygote* sur ce caractère puisqu'il porte nécessairement deux copies du gène "œil bleu". Par opposition, lorsque les deux copies d'un gène sont différentes, on parle d'hétérozygotie*. Lorsque plusieurs versions d'un gène existent, on parle de polymorphisme. Chaque version d'un gène sera appelé allèle*, et un individu avec ses caractéristiques sera appelé variant. Le polymorphisme est important dans la mesure où cela

* Voir glossaire, page 199

permet à une population d'avoir des sous-groupes aux caractéristiques distinctes. Cela permet de résister à des pathologies par exemple.

Il existe ainsi chez l'homme une mutation sur un gène qui code l'hémoglobine qui provoque une maladie appelée *drépanocytose*^{*}. Cette maladie du sang est relativement rare dans la population, car ceux qui en souffrent ont souvent une vie difficile et meurent prématurément. Mais les porteurs de cette maladie résistent mieux à l'infection paludéenne que les individus sains. Ceci a pour conséquence que dans les régions où le paludisme a une très forte incidence, finalement, les malades de drépanocytose vivent mieux et plus longtemps que le reste de la population, et dans ces régions, ce sont les porteurs de cette maladie qui sont les plus nombreux.

Une population en bonne santé doit présenter une variété de gènes différents, ce qui lui permettra de s'adapter et d'évoluer ; comme nous le verrons plus tard, cela lui permettra d'avoir une plus grande capacité de réaction vis-à-vis d'un changement dans l'environnement, ou d'une épidémie par exemple.

La dominance de certains caractères, la récessivité d'autres, la subtile régulation de l'expression des gènes, et les interactions avec l'environnement sont à l'origine d'effets qui peuvent être spectaculaires lorsqu'on regarde à l'échelle d'une population la présence de tel ou tel caractère. Si les populations sont isolées et que trop peu d'individus sont présents, il y a une sélection des caractères, et donc des gènes au sein d'une population. En effet, sur une population faible, la surreprésentation d'un gène peut aboutir à sa diffusion rapide dans la descendance et à la disparition de la ou des autres versions de ce gène, et ce sera particulièrement grave en cas de fort bouleversement de la population, comme celui lié à une épidémie.

Un cas bien connu est décrit comme effet fondateur ou colonisateur. Par exemple, lorsque les premières populations humaines ont franchi le détroit de Béring pour coloniser les Amériques, elles se sont trouvées isolées du reste du monde. Les caractéristiques génétiques qu'elles présentaient n'étaient issues que d'une petite proportion de l'humanité et elles ont donné les caractéristiques phénotypiques des amérindiens, parmi lesquelles leur système immunitaire. Lorsque les colons européens sont arrivés, ils étaient porteurs de certaines pathologies considérées comme bénignes en Europe, mais qui n'existaient pas sur le Nouveau Monde et les conséquences ont été dramatiques. Les indigènes ont été décimés par des épidémies mortelles de pathologies qui n'étaient que d'importance mineure sur le Vieux Continent.

C'est à ce même souci de variabilité génétique que sont liés aujourd'hui les problèmes de génétique des populations évoqués lorsqu'on parle des espèces en voie de disparition ou de la sélection des espèces par l'homme. Pour les espèces en voie de disparition le problème du nombre d'individus traduit le nombre de versions de chaque gène qui circulent, un polymorphisme trop faible, et une occurrence trop faible de chaque gène risque de porter un préjudice fatal si un changement brusque survient. En outre, tous les représentants de l'espèce sont rarement localisés dans une région très confinée du monde, si bien que le brassage génétique n'est pas efficace. La propagation des caractères se fait de proche en

^{*} Voir glossaire, page 199

proche, ainsi, par exemple, séparés de quelques milliers de kilomètres, les tigres du sous-continent indien s'accouplent rarement avec des tigres des îles de Java ou de Bali sans intervention humaine. De tels accouplements sont au cœur des programmes de conservation d'espèces pour essayer artificiellement d'assurer un fort brassage génétique.

Les cas de l'agriculture ou de l'élevage sont légèrement différents. L'homme a sélectionné des caractères et donc des gènes qui servaient au mieux les besoins de production. Cette sélection était faite à l'origine selon les contraintes locales, et était réalisée un peu « au petit bonheur la chance », en croisant des individus dont les caractères présentaient un fort intérêt. Avec la course aux rendements de l'agriculture intensive et la mondialisation, ce sont les mêmes caractères qui ont été sélectionnés pour répondre à un standard, quitte à aider au maximum le développement à l'aide d'engrais, hormones et autres. De plus ces caractères ont été sélectionnés à la main, précisément, à l'aide des biotechnologies. La variabilité a ainsi été très fortement réduite, et aujourd'hui, avec un changement des pratiques agricoles pour un retour à une production plus rationnelle, qui se fait moins au détriment de l'environnement, nous ne disposons plus de populations adaptées aux contraintes locales des lieux de production. Les programmes de conservation de graines ou de recherche de variantes locales ou d'espèce archaïque visent ainsi à la réintégration de caractères perdus ou l'intégration de nouveaux caractères dans les lignées destinées à l'agriculture ou à l'élevage.

Maintenant que nous avons vu dans les grandes lignes le fonctionnement de l'hérédité et les éléments fondamentaux de la génétique des populations, il nous faut nous intéresser à l'ensemble des structures qui commandent à ces règles, le génome.

II.3.2. NOTIONS DE GENOMIQUE

Nous allons nous intéresser à la structure du génome et aux différentes formes qu'il peut revêtir, son expression et comment elle est régulée, et pour finir nous parlerons de sa transmission.

II.3.2.1. STRUCTURE DU GENOME

Tous les êtres vivants, au sens de la définition que nous avons retenue, possèdent un génome structuré de la même façon. C'est ce que nous appelons aussi couramment le patrimoine génétique. Il est constitué de longues molécules d'ADN qui sont repliées sur elles-mêmes.

La molécule d'ADN est une chaîne de nucléotides*. Ces derniers sont formés de l'assemblage d'une base azotée (Adénine*, Thymine*, Guanine*, Cytosine*), d'un phosphate et d'un sucre, le désoxyribose. Le phosphate et le sucre permettent la construction de la séquence d'ADN. L'ADN est comparé à une échelle, les montants étant formés de l'enfilade de sucres et phosphates alternés, les bases azotées correspondant aux barreaux. Les nucléotides sont en général représentés par l'initiale de la base azotée qu'ils contiennent : A, T, G et C.

* Voir glossaire, page 199

Cette molécule est dédoublée de sorte qu'il y a toujours deux séquences qui sont alignées, soit en mettant deux molécules qui s'apparient, soit par une molécule unique qui se replie sur elle-même. On parle alors de séquence double-brin. C'est la forme normale et stable de l'ADN qui s'enroule sur elle-même et donne la structure en double-hélice découverte par Watson et Crick au début des années 1950.

La séquence des deux brins est complémentaire, de sorte que, connaissant la séquence de l'un, on connaît la séquence de l'autre, ceci sera important pour la reproduction et pour la réparation de dommages subis par le génome. Les règles d'appariement des bases azotées indiquent qu'en face d'un C on trouve un G, et en face d'un T, on trouve un A. Un chromosome* est constitué d'une double-hélice unique. Lorsqu'une cellule possède plusieurs chromosomes, elle possède plusieurs double-hélices d'ADN. Cette structure est à l'origine de l'unité utilisée pour exprimer la taille des génomes, la paire de base (pb).

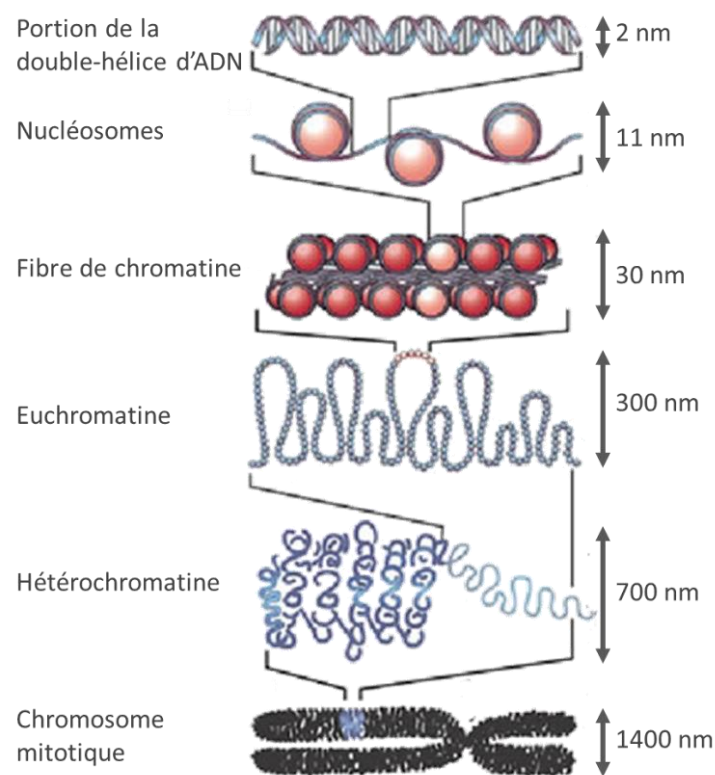


Figure II.4 : Les différentes formes de l'ADN (d'après Felsenfeld et Groudine, 2003 [1]).

La double-hélice est enroulée autour de bloc protéiques, les histones*. Deux tours autour d'une histone forment un nucléosome*. La séquence de nucléosomes s'enroule à son tour sur elle-même pour former une fibre de chromatine*. En temps normal, la chromatine est présente dans les cellules sous une forme plus ou moins condensée, l'euchromatine* ou l'hétérochromatine. Les chromosomes tels qu'on les représente souvent, en X, sont des structures très denses d'hétérochromatine* enroulée, qui ne sont présentes qu'au moment de la multiplication cellulaire (Figure II.4).

* Voir glossaire, page 199

Chez beaucoup de bactéries il existe un chromosome unique constitué d'une seule molécule d'ADN. Le chromosome est alors circulaire (Figure II.5).

Tous les êtres vivants supérieurs possèdent plusieurs chromosomes. La complexité du patrimoine génétique est liée à la longueur des séquences qui le composent et au nombre de chromosomes qui le constituent. Cette complexité n'est en rien liée à la complexité apparente de l'organisme qui le porte, le plus petit génome connu à ce jour a été identifié chez *Encephalitozoon cuniculi* (Microsporidie), il comprend 2,9 Mpb alors que le plus gros génome rencontré jusqu'à présent comprend 150 Gpb et appartient à une plante, *Paris japonica*. Pour information et pour comparaison, chez l'être humain, sa taille est de 3 Gpb.

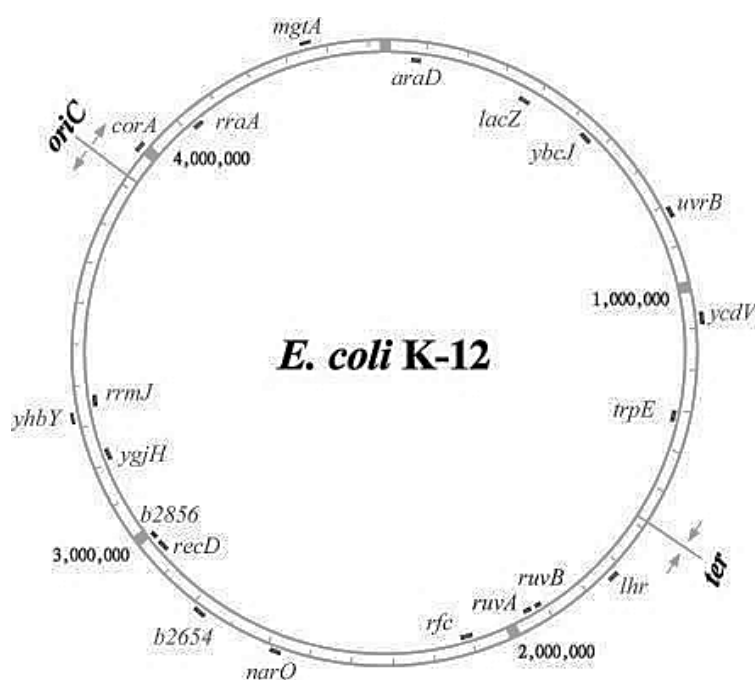


Figure II.5 : Carte génétique simplifiée d'*Escherichia coli*. L'origine de réplication (*oriC*), quelques gènes et une séquence terminatrice (*ter*) sont représentés avec leur position sur la séquence mesurée en paires de bases (Crédit Open Access Subset of PubMed Central).

II.3.2.2. EXPRESSION DU GENOME

Nous avons vu que l'ensemble des caractéristiques observables chez un individu, le phénotype, dépend de l'ensemble de gènes qu'il porte, son génotype. En connaissant le génotype on ne peut pas déterminer le phénotype précis, et inversement. L'expression du génome est complexe et subtilement régulée. Sans trop entrer dans les détails, nous nous contenterons de décrire le fonctionnement général de l'expression du génome.

La séquence des double-hélices d'ADN est lue par la machinerie cellulaire, les enzymes. Certaines portions sont dites codantes parce que leur expression permet la fabrication des protéines qui sont les effecteurs cellulaires, d'autres sont dites non-codantes. Ces séquences non-codantes peuvent soit être des séquences régulatrices, soit ne pas avoir d'utilité connue. Ce n'est pas pour autant qu'elles ne servent à rien, et si on parlait d'elles comme d'ADN poubelle jusqu'à la fin du 20^{ème} siècle, cette dénomination disparaît peu à peu du langage.

En effet, on connaît dans ces séquences non-codantes la trace d'anciens gènes, et si ceux-ci ne sont plus exprimés aujourd'hui, on suppose qu'elles servent de « réserve » pour assurer la variabilité génétique. Une autre fonction reconnue de ces séquences non codantes est une fonction de protection. Puisque l'ADN peut être la cible d'attaques chimiques, et vu l'importance d'assurer sa fonction de support de l'information génétique, la multiplication des séquences non-codantes « dilue » l'information indispensable, et permet de réduire les risques de pertes sur une séquence codante, puisqu'il y a autant de chance que ce soit une séquence non-codante qu'une séquence codante qui soit altérée.

Les séquences non-codantes, qui sont des séquences régulatrices, jouent indirectement un rôle sur l'expression du génome. Ces séquences peuvent être placées n'importe où sur la séquence d'ADN, même très éloignées de la séquence qu'elles régulent, voire même sur un chromosome différent. Par la structure qu'elles adoptent, elles peuvent empêcher l'accès à une séquence codante par la machinerie cellulaire ou au contraire stabiliser cette interaction, en calant cette machinerie au contact de la séquence codante, alors qu'elle aurait tendance à s'en décrocher.

Pour qu'une séquence codante s'exprime, il faut qu'elle soit transcrite en séquence ARN*. C'est le rôle de la machinerie cellulaire de lire la séquence d'ADN et de synthétiser la séquence ARN correspondante. L'ARN répond à la même règle d'appariement que l'ADN et une séquence d'ARN est donc complémentaire de la séquence d'ADN qui sert de matrice à sa fabrication. Considérons deux différences notables toutefois : il n'y a pas de T dans l'ARN, ils sont remplacés par des U (de la base azotée Uracile qui se substitue à la Thymine), et le sucre n'est pas du Désoxyribose mais du Ribose.

La séquence d'ARN est traitée ensuite par les ribosomes pour réaliser la synthèse des protéines. Lorsqu'on parle de machinerie cellulaire, ce sont presque toujours des protéines particulières, les enzymes, dont on parle. Il existe d'autres molécules actives au sein d'une cellule, mais, pour simplifier la description du système, nous n'en parlerons pas.

Les séquences codantes ne sont pas toujours toutes exprimées dans toutes les cellules. Une cellule peut éteindre certaines séquences pour se spécialiser, au sein d'un organe par exemple, ou activer certaines séquences pour réagir à une situation nouvelle. Cette inhibition peut se faire soit par des moyens mécaniques, la portion d'ADN qui porte la séquence est enroulée sous forme d'hétérochromatine ou est coincée dans une gaine d'hétérochromatine, soit par des moyens chimiques. Certaines bases sont alors modifiées pour ne plus être reconnues par les enzymes, ou d'autres molécules sont associées à la séquence pour empêcher son accès par les protéines. Notons que parfois, ce qui est vu comme une activation d'un gène est en fait une levée d'inhibition, c'est-à-dire que le gène est éteint mécaniquement ou chimiquement en temps normal, et qu'au besoin, la cellule lève le verrou.

Chez les organismes qui pratiquent la reproduction sexuée il y a une régulation globale qui s'exerce sur l'ensemble du génome. Cette régulation très complexe n'a commencé à être décrite que depuis le début des années 2000, c'est l'épigénétique. Les individus appartenant à des espèces chez qui la reproduction sexuée a cours sont porteurs d'un jeu de chromosomes hérité du père et d'un jeu de chromosomes hérité de la mère. Avec les

* Voir glossaire, page 199

premières expériences de clonage, on s'est aperçu que certains gènes qui s'expriment sont toujours issus du patrimoine maternel. Un gène du patrimoine paternel peut être exprimé en cas de problème avec la copie issue de la mère. D'après les règles établies par Mendel, il devrait y avoir codominance, mais ce n'est pas le cas, on n'observe l'expression que d'une version du gène. Le phénotype obtenu est différent de celui attendu à la lecture du génotype et inversement, le génotype ne sera pas tout à fait celui déduit de l'observation d'un phénotype.

Il semble que l'inhibition incriminée soit également chimique et/ou mécanique, comme la régulation normale, mais les mécanismes qui contrôlent l'épigénétique sont encore mal connus. On est par contre sûr que l'environnement joue un rôle important dans ce phénomène, et beaucoup de polluants et perturbateurs endocriniens ont vu leur effet compris grâce à l'épigénétique.

L'épigénétique est également le seul phénomène connu qui modifie l'expression d'un ou plusieurs gènes qui puisse être transmis à la descendance. On a ainsi remarqué que les femmes des pays d'Europe du nord qui avaient subi des privations et des carences pendant la seconde guerre mondiale avaient des enfants qui présentaient plus de risque de développer du diabète que les autres. En ayant été sous-alimentées, elles ont subi une modification des gènes codant les acteurs de la régulation de la glycémie, et avec le retour à une situation faste, cette modification a été annulée. Pourtant les lignées germinales en ont gardé la trace et leurs enfants ont hérité d'un patrimoine « stressé » qui peine à gérer normalement la glycémie.

La régulation de l'expression génétique est complexe, et nous venons de le voir, elle est fortement liée à la transmission du patrimoine génétique entre les parents et les enfants, aussi nous allons maintenant détailler un peu plus cette transmission.

II.3.2.3. TRANSMISSION DU GENOME

La transmission du patrimoine à l'enfant est dite verticale. Un nouvel individu possède forcément une version complète de son génome lorsqu'il est formé. Chez les microorganismes unicellulaires, lors de la division de la cellule mère, chaque cellule fille reçoit une copie du génome. Chez les organismes qui ont recours à la reproduction sexuée, le patrimoine génétique est hérité pour moitié de chacun des parents. Avec ce que nous avons décrit jusqu'à présent, nous avons surtout considéré ce dernier mode de reproduction, plus facilement observable.

La reproduction sexuée fait intervenir les cellules sexuelles, généralement produites dans des organes spécialisés, les gonades. Leur production fait intervenir un processus complexe, la méiose*, qui permet, à partir d'une cellule, de produire quatre gamètes porteurs chacun d'une moitié du patrimoine génétique de l'individu. Lors de la méiose, les chromosomes sont dupliqués et puisque les chromosomes vont par paire (une version maternelle et une version paternelle), une fois dupliqués, il y a quatre versions de chaque chromosome dans la cellule. Sommairement, celle-ci se divise alors à deux reprises, et à chaque fois les cellules filles ne

* Voir glossaire, page 199

prennent que la moitié des chromosomes. Au final, chaque gamète ainsi produit a obtenu une version de chaque chromosome.

Au contraire de la reproduction asexuée, la reproduction sexuée permet un brassage génétique efficace. En effet, au cours de la méiose, les chromosomes issus du père se mélangent à ceux issus de la mère. Chez les êtres humains, par exemple, avec 23 paires de chromosomes, il est rare de trouver chez un individu un gamète qui n'aura reçu que les chromosomes issus de sa mère par exemple. De plus, lors de la méiose, il peut arriver des réarrangements chromosomiques, les chromosomes d'une paire échangent des séquences, de sorte qu'un chromosome n'est plus exclusivement celui hérité d'un des parents, mais est un mélange des deux, c'est le *crossing-over*.

Certains microorganismes sont capables de se reproduire de façon sexuée, et en général, ils réservent à ce mode un rôle de réponse à un stress. En effet, la reproduction sexuée permet un plus grand brassage génétique, et nous avons vu qu'il est important pour la bonne santé d'une population. Comme la reproduction asexuée ne consiste qu'en la duplication du patrimoine génétique, puis à sa répartition en quantité égale entre les cellules filles, seules les mutations peuvent apporter du changement dans le patrimoine.

Notons une précision importante, la transmission du patrimoine génétique que nous avons évoquée jusqu'à maintenant concerne la transmission d'une génération à l'autre. Il faut aussi considérer que le patrimoine génétique est transmis à chaque cellule d'un organisme complexe. Nous avons dit que celui-ci est créé à la fécondation par la rencontre d'un gamète mâle et d'un gamète femelle ; l'œuf qui est alors obtenu se divise pour former toutes les cellules qui composent l'organisme, et chacune dispose de son propre patrimoine génétique. À chaque division cellulaire, le génome est dupliqué, et chaque cellule fille en hérite d'une copie, identique si ce n'est l'apparition éventuelle de mutations. De façon triviale, on peut considérer que les cellules d'un organisme supérieur utilisent la reproduction asexuée pour former toutes les cellules dudit organisme. Ce mode de division s'appelle la mitose*, dont nous reparlerons un peu plus loin (page 43). Les mécanismes de régulation de l'expression des gènes permettent ensuite aux cellules de se différencier en cellules spécialisées.

Avant de regarder plus en détail la structure des cellules, prenons le temps de mentionner une spécificité des organismes unicellulaires. Il existe chez eux des séquences d'ADN indépendantes du génome. Ces séquences forment une structure circulaire autonome, qui peut être répliquée indépendamment des chromosomes, de sorte que l'on peut en trouver plusieurs exemplaires dans une seule cellule. Ces structures sont appelées plasmides*. Si la transmission de parent à enfant du patrimoine génétique est appelée transmission verticale, par opposition la transmission de plasmide peut être horizontale, elle peut se faire de cellule à cellule, par contact direct, en plus de pouvoir se faire de façon verticale. Chez une même cellule, on peut trouver plusieurs plasmides différents, chacun en un ou plusieurs exemplaires. Les plasmides ne portent jamais de gènes essentiels ni caractéristiques d'une espèce.

* Voir glossaire, page 199

II.4. L'UNITE FONCTIONNELLE DU VIVANT : LA CELLULE

II.4.1. STRUCTURE DES CELLULES

Nous avons mentionné l'existence de cellules dites procaryotes et de cellules dites eucaryotes, respectivement les cellules dépourvues d'un noyau et les cellules dotées d'un noyau. Si pour l'instant nous n'avons vu qu'une différence au niveau de la transmission du patrimoine génétique, nous allons maintenant décrire ces cellules de façon un peu plus détaillée, en commençant par ce qui est commun à toutes les cellules, avant d'essayer de préciser les différences qui existent entre procaryotes et eucaryotes.

II.4.1.1. ASPECTS GENERAUX

Une cellule, pour être autonome et « différente » de ce qui l'entoure, doit être délimitée par une enveloppe bien définie, au minimum une seule enveloppe, la membrane cellulaire ou membrane cytoplasmique, qui sépare l'intérieur de la cellule, nous parlerons de milieu intracellulaire (MIC), de l'extérieur, nous parlerons de milieu extracellulaire (MEC). Le milieu intracellulaire est une solution saline appelée cytoplasme* ou cytosol* qui contient toutes les molécules constitutives d'une cellule et dans lequel baignent toutes les structures cellulaires.

Chaque cellule dispose de son patrimoine génétique propre, avec son ou ses chromosomes, et éventuellement, son ou ses plasmides. Parmi les molécules qui constituent la cellule, on trouve toute la machinerie qui permet à la cellule de fonctionner, c'est-à-dire de répliquer le génome, d'exprimer les gènes, et donc de produire des protéines qui sont véritablement les effecteurs cellulaires assurant la quasi-totalité des fonctions. Parmi ces fonctions, citons la structure, le transport, la régulation, le contrôle, la synthèse, la transformation et le recyclage, la production d'énergie, etc... Elles sont représentées à la Figure II.6.

Les protéines de structure constituent son ossature lorsqu'elles sont à l'intérieur de la cellule, et lorsqu'elles sont à l'extérieur de la cellule, elles forment une paroi solide. Ce sont elles qui confèrent sa forme à une cellule. Elles sont également responsables de la mobilité de la cellule, par exemple des cils ou des flagelles qui permettent à une cellule de « nager », ou des déformations transitoires de la cellule et la formation de « pieds », qui vont permettre un déplacement par une forme de reptation, on parlera alors d'haptotaxie*.

Les protéines de transport permettent les échanges de matière entre la cellule et son environnement. Elle va pouvoir absorber des composés, (nutriments, facteurs de croissance, ou éléments toxiques, etc...) et en rejeter d'autres (déchets, enzymes externes qui vont aider à utiliser les composés du milieu, antibiotiques, etc...). Leur rôle est également important pour la communication entre les cellules, en autorisant le transit de messages chimiques, ou assurant la détection d'un signal chimique. Elles sont également impliquées dans les échanges entre les compartiments cellulaires.

* Voir glossaire, page 199

Les protéines de contrôle et de régulation permettent d'accélérer ou de ralentir certaines fonctions cellulaires, les processus de synthèse par exemple, ou d'assurer la conduction des messages intracellulaires, ou encore de vérifier l'intégrité des constituants cellulaires. Elles participent aussi à l'adressage des composants synthétisés vers l'endroit où ils sont destinés, à la séquestration de composés dans certains compartiments, au blocage de l'accès de composés à certains compartiments.

Les protéines de transformation ou de recyclage sont impliquées dans des fonctions que l'on regroupe sous le terme de catabolisme* sur lequel nous reviendrons un peu plus tard (voir page 47). Ce sont elles qui vont permettre de dégrader les nutriments pour en récupérer les matériaux utilisables dans les processus de synthèse. Elles permettent aussi de recycler le matériel mal formé ou abîmé. Ces protéines peuvent être libérées dans le milieu extracellulaire pour exploiter certains de ses composés.

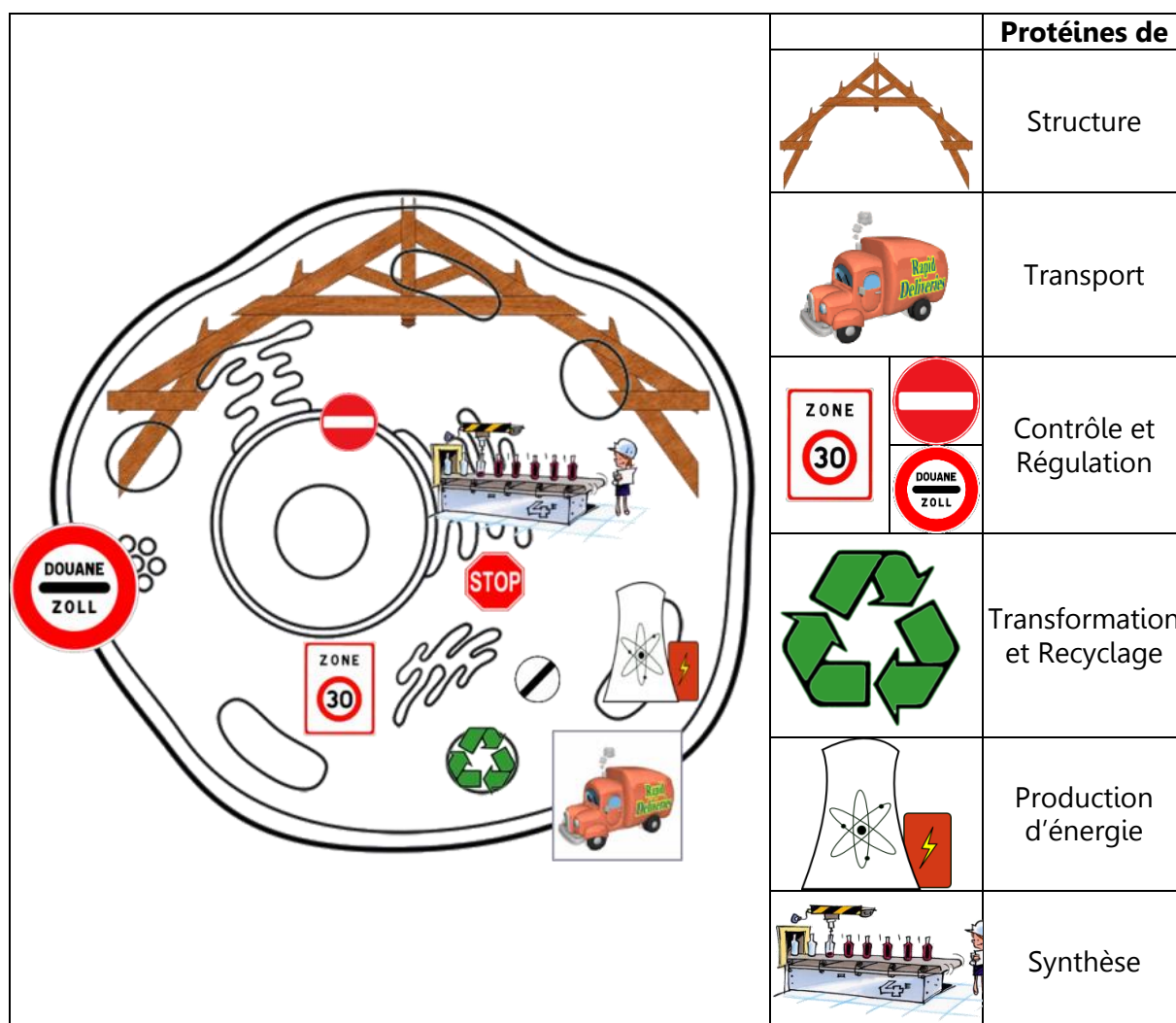


Figure II.6 : Schéma d'une cellule eucaryote et de ses principales fonctions.

Les protéines de production d'énergie permettent d'extraire l'énergie contenue dans les nutriments pour produire l'énergie nécessaire à la cellule, elles aussi font partie intégrante du

* Voir glossaire, page 199

catabolisme. Cette énergie peut être exploitée sous forme de chaleur (au niveau des muscles, elle permet la thermogenèse* chez les animaux homéothermes, par exemple), ou sous forme d'énergie chimique ou mécanique, qui va pouvoir être utilisée dans les réactions du métabolisme* ou par les protéines de structure pour produire du mouvement.

Les protéines de synthèse sont impliquées dans tous les processus de production et de fabrication du matériel cellulaire ; toutes ces fonctions sont regroupées sous le terme d'anabolisme* qui sera abordé un peu plus loin dans ce chapitre (page 47). La synthèse peut également concerner des composés qui seront rejetés à l'extérieur de la cellule, soit pour faciliter l'exploitation du milieu, ce sera le rôle des enzymes, soit pour limiter la croissance d'organismes concurrents, comme le permettront des antibiotiques, par exemple.

Toutes les protéines sont constituées d'acides aminés et ont la même origine. C'est-à-dire, une ou plusieurs séquences d'ADN ; soit un ou plusieurs gènes, transcrites en séquences ARN, traduites à leur tour en séquences peptidiques, selon les règles du code génétique. La traduction est l'œuvre de structures spécialisées appelées ribosomes*, présentes en grand nombre dans le cytoplasme. Cette séquence peptidique va se replier pour former la protéine fonctionnelle, ou éventuellement, une sous unité de la protéine fonctionnelle. Le cas échéant, les sous unités s'assemblent pour former la protéine fonctionnelle. Une étape de maturation supplémentaire peut être nécessaire pour adjoindre à la protéine des groupements chimiques annexes et la rendre pleinement fonctionnelle.

Les nucléotides de la séquence d'ARN sont lus 3 par 3. Le code génétique, universel chez tous les êtres vivants sur Terre, associe un acide aminé à chaque triplet de nucléotides. Pour chaque triplet lu, le ribosome ajoute l'acide aminé correspondant à la séquence peptidique.

Il peut exister des compartiments destinés au stockage au sein d'une cellule, on les appelle vacuoles*. Leur présence est systématique chez les cellules de certaines espèces. Le cytoplasme peut également contenir des grains d'inclusions. Ces grains sont souvent des réserves de nutriments ou des stocks de sels. Comme les vacuoles, leur présence dépend souvent de l'état nutritionnel des cellules ou des conditions environnementales.

À l'exception de compartiments dédiés au stockage qui seront des vésicules ou des vacuoles, lorsque plusieurs fonctions spécialisées similaires sont regroupées au sein d'une structure spécialisée, on parlera d'organites* ou d'organelles*. À l'exception des ribosomes, il n'existe pas d'organites chez les procaryotes.

II.4.1.2. PRESENTATION DE LA CELLULE PROCARYOTE

Procaryote, du grec *Karuo*, le noyau, et pro, primitif, désigne des organismes unicellulaires dont les cellules ont une structure simple, dépourvues d'un noyau figuré. Ces organismes, qui regroupent les eubactéries et les archaebactéries, ont leur patrimoine génétique libre dans le cytoplasme, et ne possèdent pas d'organite compartimenté, mais certaines espèces peuvent toutefois présenter une vacuole, de taille plus ou moins importante.

* Voir glossaire, page 199

À l'exception des mycoplasmes, l'enveloppe des cellules procaryotes est plus élaborée qu'une simple membrane, lui associant d'autres structures dont la plus fréquente est une paroi. La paroi bactérienne sert d'élément de classification des bactéries, on trouve en effet deux grandes familles : les bactéries disposant d'une paroi épaisse, dites à gram positives, et celles disposant d'une paroi plus fine couverte par une autre membrane, qui sont dites à gram négatives.

La forme et la taille des organismes procaryotes sont très variables, généralement sphéroïdales, ou en forme de petit bâtonnets, (Figure II.7), de dimension proche de 2 μm , elles peuvent aller de 0,1 μm de diamètre pour les espèces de bactéries les plus petites (*Chlamydiae*, *Hemobartonelles*) à 750 μm de diamètre pour les bactéries les plus imposantes (*Thiomargarita namibensis*).

Le génome de ces organismes peut présenter une très grande variété de forme et d'organisation. La plupart des bactéries possèdent un chromosome unique et circulaire qui se concentre au centre de la cellule, en ce que l'on nomme un nucléoïde*. La taille du génome bactérien peut aller de 160 kpb pour les bactéries dotées des plus petits génomes (*Candidatus Carsonella ruddii*) à 12,2 Mpb pour celles au génome le plus fourni (*Sorangium cellulosum*).

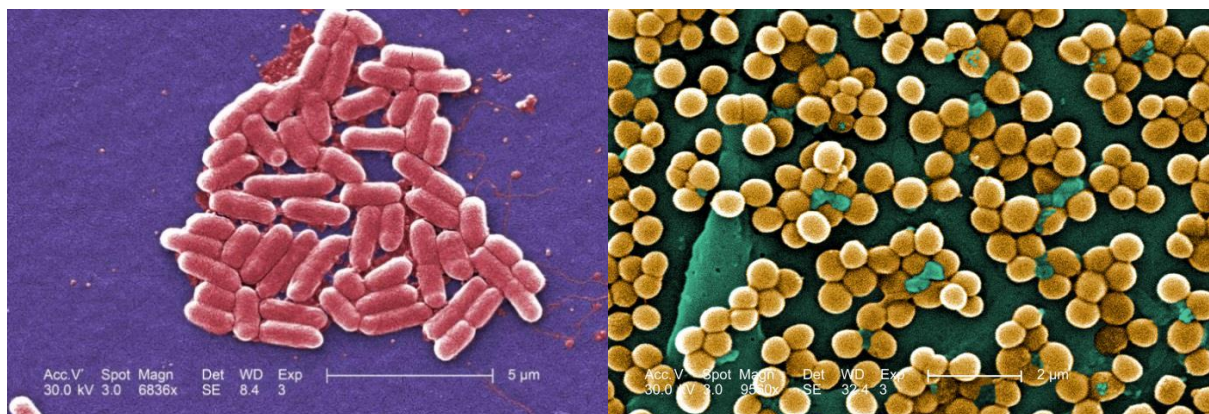


Figure II.7 : A gauche : *Escherichia coli* ; à droite : *Staphylococcus aureus*. Images en fausses couleurs obtenues en microscopie électronique à balayage (x10000) ; crédit Public Health Image Library.

La mobilité des procaryotes est assurée par des flagelles et/ou des cils. Leur disposition est souvent caractéristique de l'espèce. Le mouvement de ces cils et flagelles, et leur disposition sur la cellule détermine le mouvement global des cellules. Ainsi, une ciliature dite monotriche*, ou polaire, présentant un seul flagelle, favorisera un déplacement en ligne droite, alors qu'une ciliature dite péritriche* avec plusieurs flagelles répartis sur toute la surface cellulaire, favorisera un déplacement à l'apparence erratique.

II.4.1.3. PRESENTATION DE LA CELLULE EUCARYOTE

Les organismes eucaryotes présentent des structures variées, selon qu'ils appartiennent à des organismes unicellulaires ou pluricellulaires, et selon qu'ils appartiennent à l'ordre des protistes, des végétaux, des champignons ou des animaux.

* Voir glossaire, page 199

Tous ont en commun de présenter un noyau délimité par une double membrane, et plusieurs organites compartimentés : réticulum endoplasmique*, appareil de golgi*, mitochondries*... Le matériel génétique, à l'exception des plasmides éventuels, est séquestré dans le noyau. Leur cytoplasme contient également des structures bien définies mais non compartimentées, comme les centrosomes*, et un très grand nombre de ribosomes. Il peut également contenir des grains d'inclusions qui servent de réserve, ou qui séquestrent des composés nocifs, ou des sels.

Le génome peut être constitué d'un nombre de chromosomes compris entre un et plusieurs dizaines, et sa taille est très variable, de 2,9 Mpb chez *Encephalitozoon cuniculi* (Microsporidie) à 150 Gpb chez *Paris japonica*.

La production d'énergie est dévolue à un organe spécifique, les mitochondries.

Les protistes ne représentent que des organismes unicellulaires dont la mobilité est assurée par des cils, des flagelles ou par haptotaxie.

Les végétaux regroupent les algues et les plantes. Certaines algues sont unicellulaires. Les cellules végétales sont caractérisées par la présence de chloroplastes*, qui sont le siège de l'activité photosynthétique. Elles contiennent une grande vacuole de réserve, qui peut occuper la quasi-totalité du volume cellulaire, et qui sert de stockage pour de l'eau et des solutés variés, comme des sucres, des ions, des acides aminés... Les cellules végétales ont une forme anguleuse définie par leur épaisse et rigide paroi cellulosique chez les végétaux complexes. Les algues microscopiques élaborent une paroi dont la composition chimique dépend de l'espèce, généralement du carbonate de calcium, ou de la silice. Ces algues peuvent être dotées de flagelles pour assurer leur mobilité.

Parmi les champignons, on trouve des organismes unicellulaires, représentés par les levures et champignons filamenteux, et des organismes pluricellulaires, représentés par les champignons dits supérieurs, ou eumycètes*. Ce que l'on appelle couramment champignon n'est qu'une structure reproductrice transitoire des eumycètes, le sporophore*. Les cellules fongiques présentent également une variété de formes, de la forme ovoïde des levures à la forme cylindrique des cellules des champignons filamenteux, et elles possèdent une paroi chitineuse.

Les cellules animales ne sont présentes que chez des organismes pluricellulaires et ce sont celles qui présentent la plus grande variété de formes et de tailles. De quelques micromètres à plusieurs mètres de long, d'une forme sphéroïdale à une forme étoilée. Les cellules animales sont dépourvues de paroi, mais, pour la plupart, elles s'entourent d'une couche flexible qui leur permet d'adhérer à leur support. Cette couche s'appelle le glycocalix*.

Pour subsister, un être vivant puise dans son environnement de la matière pour produire de nouveaux constituants ou les renouveler, et de l'énergie pour assurer son bon fonctionnement. « Rien ne se perd, rien ne se crée, tout se transforme » (Lavoisier), si une cellule capte dans son environnement des nutriments, comme elle ne grossit pas

* Voir glossaire, page 199

indéfiniment, leur utilisation ou le renouvellement des constituants cellulaires s'accompagne de la production de déchets, que la cellule devra rejeter dans son milieu.

II.4.2. VUE D'ENSEMBLE DE LA VIE DES CELLULES

Depuis le 19^{ème} siècle et les travaux de Pasteur, on sait que les cellules ne peuvent venir que de cellules existantes. Il existe donc un moyen pour les cellules de se multiplier. Ce moyen, c'est la division cellulaire, qui voit deux cellules se former à partir d'une cellule. Chez les organismes unicellulaires, la division cellulaire est directement responsable de leur prolifération dans leur environnement, alors que chez les organismes pluricellulaires, la division cellulaire assure la création de nouvelles cellules. Dans ce cas elle se poursuit tout au long de la vie de l'organisme et son rôle sera d'assurer sa croissance ou de mettre à disposition de nouvelles cellules, pour remplacer les cellules vieillissantes ou mortes. Chez l'être humain, on estime que 25 millions de cellules se divisent à chaque seconde, et par an on estime que cela représente chez un adulte son propre poids en cellules qui est ainsi renouvelé. Quel que soit l'organisme, unicellulaire ou pluricellulaire, l'environnement est limité ou limitant, et les cellules se renouvellent, de jeunes cellules prenant la place de cellules mortes.

La cellule qui se divise est appelée cellule mère, les cellules qui en résultent sont appelées cellules filles. Les cellules filles sont des copies de la cellule mère. Le cycle de vie d'une cellule s'articule généralement autour de trois phases : la phase de croissance, de division, et la mort. La durée de ce cycle est très variable selon le type de cellule et selon l'espèce à laquelle elle appartient. Cette durée est en plus très fortement influencée par l'environnement qui baigne la cellule.

II.4.2.1. CROISSANCE

Le terme de croissance ne signifie pas forcément ici que la cellule grandit. Certaines espèces vont effectivement avoir une croissance au sens d'une cellule qui voit sa taille augmenter, mais chez d'autres espèces, une cellule gardera la même taille tout au long de sa vie, n'augmentant son volume qu'au moment de se préparer à se diviser.

Cette phase correspond à l'essentiel de la vie de la cellule, elle produit et entretient son matériel à partir des ressources (chimiques et énergétiques) disponibles. Cela lui permet d'assurer son fonctionnement normal, de s'adapter à une modification de son environnement, et éventuellement de « grandir » en prélude à une division. Cette phase dure de quelques minutes à plusieurs semaines, selon l'espèce et le type de cellule. Lorsque la cellule se prépare à se diviser, elle duplique tous ses constituants de sorte que chaque cellule fille puisse hériter d'une copie. Chez certaines espèces la cellule double alors de volume, ainsi chaque cellule fille a déjà son volume « adulte » lors de sa formation.

Chez les eucaryotes, le génome est concerné par cette phase et il est dupliqué à la fin de la croissance, alors que chez les procaryotes le génome ne sera dupliqué qu'à l'étape suivante, en parallèle à la division.

II.4.2.2. DIVISION

La division cellulaire est le moyen courant le plus élémentaire de multiplication des cellules. Sauf pour certains modèles de division spécifiques chez les eucaryotes, la division donne deux cellules filles identiques, qui se répartissent les constituants de la cellule mère équitablement. Chaque cellule fille est ainsi en mesure de mener sa propre existence autonome et va effectuer sa phase de croissance, synthétiser ce qui lui manque et éventuellement atteindre, si besoin, son volume « adulte », pour poursuivre le cycle.

Chez les procaryotes la division porte le nom de scissiparité*, alors qu'on préférera le terme de mitose chez les eucaryotes. Chez les procaryotes, c'est en parallèle à l'allongement de la cellule qui accompagne la division que se déroule la duplication du patrimoine génétique, alors qu'elle a eu lieu au cours de la croissance, chez les eucaryotes.

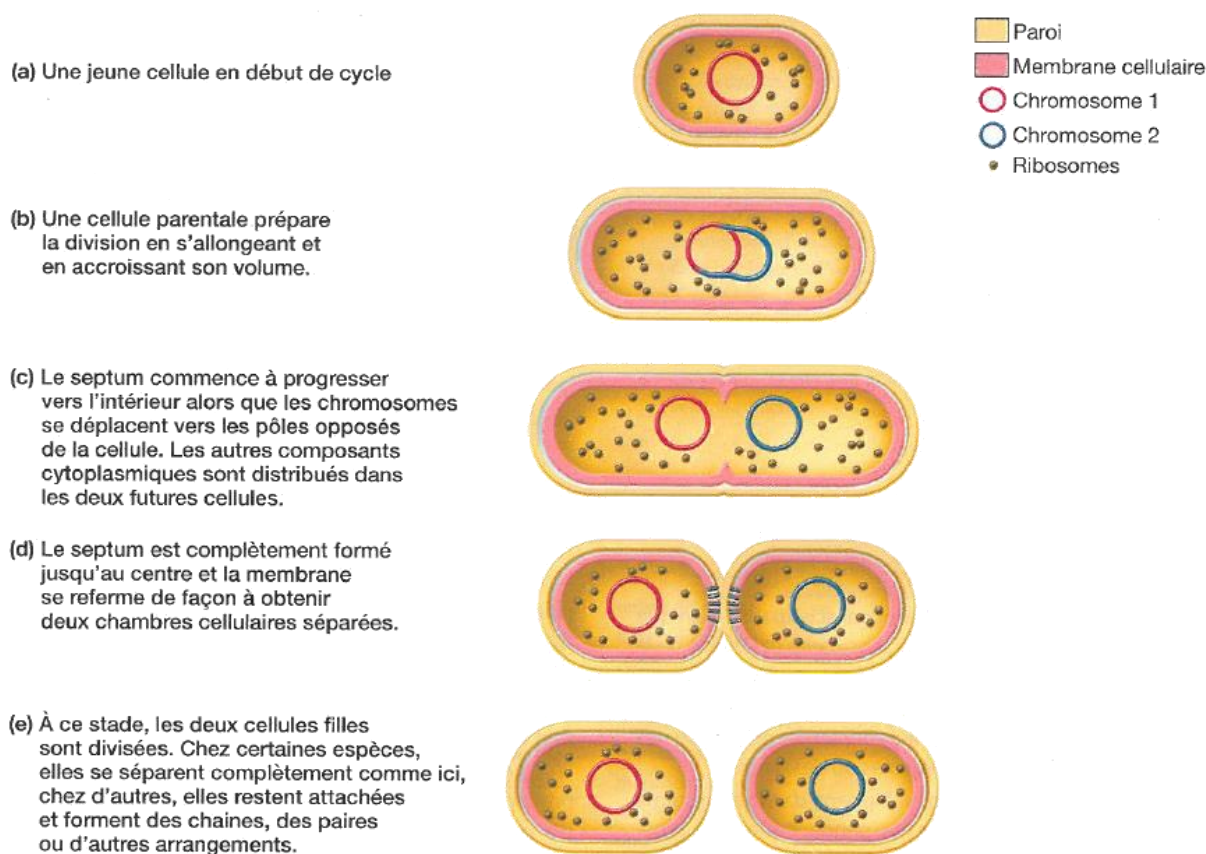


Figure II.8 : Schéma de la division cellulaire chez les procaryotes, la scissiparité (Prescott et al., 2010 [2]).

Le processus de division proprement dit se met en place lorsque la cellule atteint un volume critique. Il s'agit là d'une généralité qui ne dépend pas de la morphologie de la cellule. Elle se passe toujours de la même manière, par la mise en place d'un anneau qui se contracte ou la construction d'une cloison au niveau de la moitié de la cellule mère. Le développement de l'enveloppe de la cellule mère peut se faire par ses extrémités, ou par ses flancs et lorsque la

* Voir glossaire, page 199

division est terminée, soit les cellules filles se séparent complètement, soit elles restent liées selon les espèces (Figure II.8).

II.4.2.3. MORT

La mort de la cellule se traduit par un arrêt de ses fonctions. Elle peut être accidentelle, liée à un dysfonctionnement de la machinerie cellulaire, liée à une cause extérieure, comme la présence d'un élément toxique par exemple, ou elle peut être programmée. Ce dernier cas est une réponse de la cellule à un stress, ou à un signal endogène (par exemple, en cas d'infection virale, la cellule qui se sait infectée décide « de se sacrifier » pour empêcher le virus de proliférer) ou exogène (une cellule qui devient cancéreuse est repérée par le système immunitaire qui « lui donne l'ordre » de se sacrifier). Chez les eucaryotes, lorsque la mort est programmée, on parle d'apoptose*.

La matière qui la constitue est alors rendue à l'environnement et pourra être utilisée par d'autres cellules. Les constituants de la cellule se dégraderont spontanément dans le temps ou seront dégradés par des enzymes, issues de la cellule elle-même, ou issues de cellules voisines, pour être recyclés et réutilisés.

II.4.3. VUE D'ENSEMBLE DU FONCTIONNEMENT CELLULAIRE

II.4.3.1. METABOLISME

Le métabolisme est la somme du catabolisme, c'est-à-dire l'ensemble des processus de dégradation à l'œuvre dans la cellule, et de l'anabolisme, c'est-à-dire l'ensemble des processus de synthèse.

Tous ces processus biochimiques répondent aux lois de la chimie et de la thermodynamique. Ces processus impliquent un grand nombre de réactions chimiques.

Les réactions chimiques au sein du vivant se déroulent dans un environnement qui est parfois défavorable, ou au sein duquel la réaction serait normalement très lente. Pour faciliter ces réactions ou pour les rendre possibles le vivant utilise des catalyseurs. Les catalyseurs biochimiques sont représentés par une classe particulière de protéines, les enzymes. Comme tous les catalyseurs, les enzymes sont retrouvées intactes à la fin de la réaction chimique qu'elles ont favorisée.

Cependant, leur structure complexe les rend fragiles, il faudra donc qu'une cellule puisse les renouveler. Les enzymes, comme toutes les protéines, sont constituées d'une chaîne linéaire d'acides aminés, synthétisée par les ribosomes, qui est ensuite repliée sur elle-même.

Les réactions biochimiques dépendent de la température. Une température élevée favorise une grande mobilité des molécules, qui ont ainsi plus de chance d'entrer en interaction. De plus, lors de collisions entre les molécules, l'énergie sera plus importante si les molécules

* Voir glossaire, page 199

sont rapides. Une grande énergie sera parfois nécessaire pour qu'une réaction puisse avoir lieu, et il faudra pourvoir à un apport supplémentaire en énergie. Certaines réactions ne pourront ainsi pas avoir lieu si la température est trop basse.

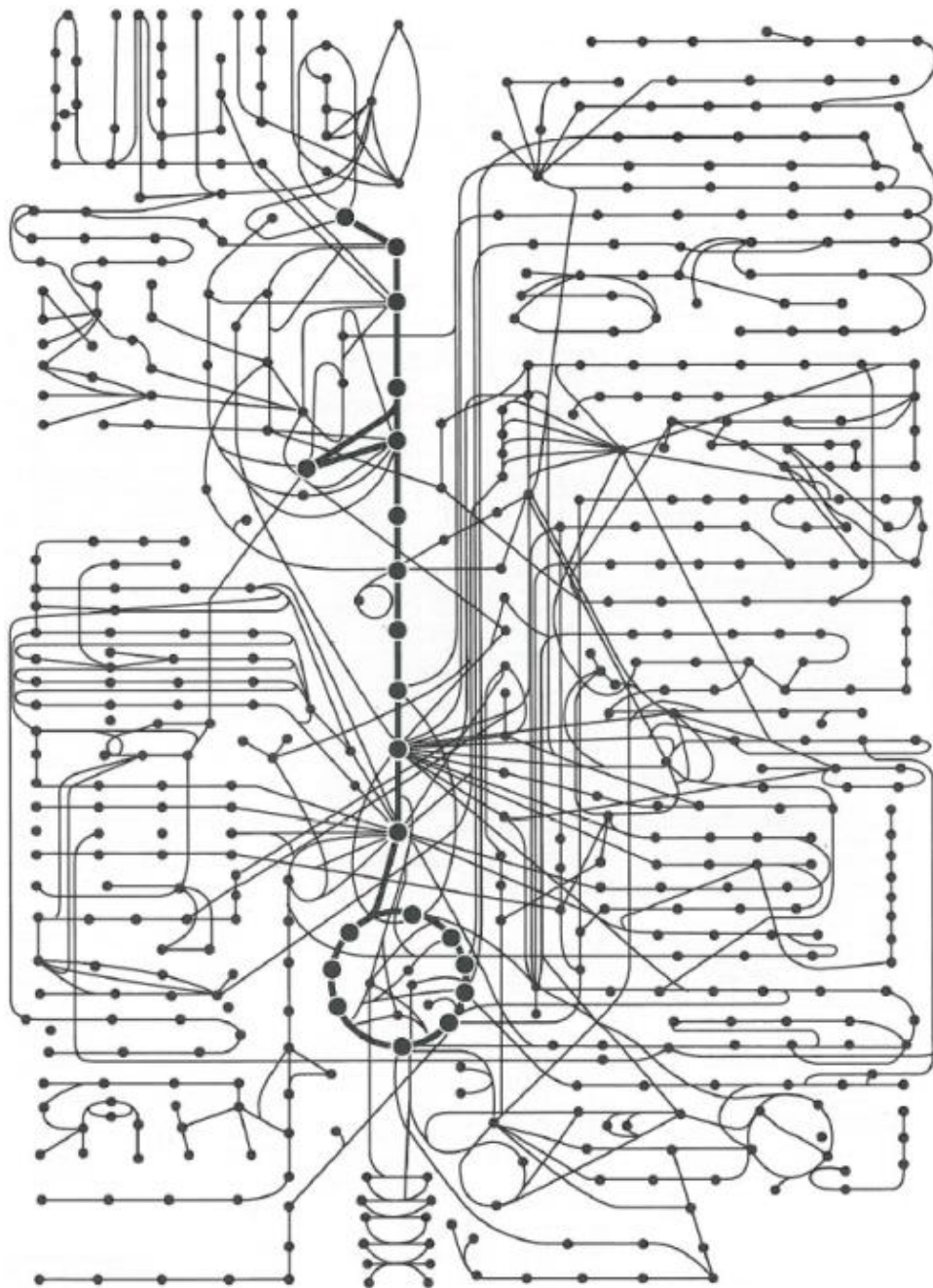


Figure II.9 : Schéma de la carte du réseau métabolique (Garret et Grisham, 2000 [3]).

Si la température est trop basse, l'efficacité des enzymes est limitée, alors que si la température est trop haute, l'agitation moléculaire aura tendance à dénaturer l'édifice protéique. Les enzymes ont ainsi une température à laquelle leur fonctionnement est optimal. La température qui correspond le mieux au métabolisme d'un type de cellule donné correspond à la température où le développement de la cellule est optimal. Les mammifères et les oiseaux ont trouvé en maintenant leur température interne à un niveau constant une

solution leur permettant d'avoir toujours un métabolisme optimal en s'affranchissant de la température extérieure qui est une contrainte sur le fonctionnement des autres êtres vivants.

Anabolisme et catabolisme sont simultanés dans la cellule, et sont souvent couplés de sorte que l'énergie libérée par les réactions cataboliques soit mise à disposition et exploitée par les réactions anaboliques dans le même temps. Certaines des réactions sont communes, de même que certains composés interviennent aussi bien dans l'anabolisme que dans le catabolisme. Les séquences de réactions forment des voies métaboliques. Plusieurs voies métaboliques interconnectées forment un réseau métabolique (Figure II.9).

Anabolisme et catabolisme sont par contre régulés de façon distincte et stricte pour que les deux n'interfèrent pas, par exemple il serait inutile de commencer à dégrader une protéine alors qu'elle n'est pas encore achevée. Pour cela, les réactions biochimiques appartenant à l'anabolisme ou au catabolisme sont souvent localisées précisément dans la cellule, et chez les eucaryotes, dans des compartiments spécialisés.

II.4.3.1.1. CATABOLISME

Le catabolisme concerne aussi bien la dégradation des nutriments captés dans l'environnement, la dégradation de composants de l'environnement afin d'en faciliter l'assimilation, ou encore la dégradation des constituants cellulaires altérés pour les recycler. Son but est de produire de l'énergie et des « briques » élémentaires, qui seront utilisées dans l'anabolisme. L'énergie peut être produite sous forme chimique, thermique, électrochimique ou mécanique.

L'énergie chimique pourra être stockée dans des molécules, dans des liaisons chimiques particulières. La molécule la plus représentative des stocks d'énergie chimique dans les cellules est l'ATP, Adénosine Tri-Phosphate. La molécule d'ATP contient deux liaisons chimiques qui libèrent beaucoup d'énergie lorsqu'elles sont cassées. La première libèrera de l'ADP et du phosphate, puis la seconde produira de l'AMP et du phosphate. Une cellule en pleine forme contient beaucoup plus d'ATP que d'ADP ou d'AMP. Le ratio entre ATP et ADP est un bon indicateur de l'état métabolique d'une cellule.

L'énergie thermique permet d'accélérer le mouvement des molécules, ou d'activer certaines réactions chimiques. Cette énergie thermique est très importante pour les animaux dits homéothermes, oiseaux et mammifères, qui maintiennent leur température interne à un niveau constant. L'énergie mécanique permet de produire du mouvement, qui sera utilisé pour déformer les protéines de structure et permettre le déplacement par haptotaxie, ou pour permettre aux cellules de se contracter, ou pour actionner les flagelles.

II.4.3.1.2. ANABOLISME

L'anabolisme regroupe toutes les fonctions de synthèse, qu'elles concernent la fabrication des constituants cellulaires, ou la production de composés destinés à être libérés dans l'environnement, ou encore l'élaboration de « briques » élémentaires ou de nutriments destinés à être stockés.

La fabrication des constituants cellulaires va largement impliquer les ribosomes, qui sont à l'origine de la synthèse des protéines. Toutes les fonctions cellulaires seront ainsi concernées.

Les composés libérés dans l'environnement peuvent être impliqués dans le catabolisme, servir à modifier l'environnement, ou constituer des messages chimiques adressés à d'autres cellules. Dans le premier cas, ce seront par exemple, des enzymes dont le rôle est de dégrader des constituants du milieu et d'en extraire des nutriments. Dans le second cas, par exemple la production d'antibiotiques permettra de rendre l'environnement néfaste pour des organismes concurrents. Les hormones représentent un bon exemple du dernier cas, en permettant la communication entre cellules.

L'anabolisme consomme les « briques » élémentaires libérées par le catabolisme, ainsi que l'énergie, qu'elle soit sous forme thermique, chimique, mécanique ou électrochimique.

II.4.3.2. ÉCHANGES ENTRE LA CELLULE ET SON ENVIRONNEMENT, NOTIONS DE TRANSPORT

Les échanges de matière entre la cellule et son environnement doivent se faire à travers son enveloppe, ce qui implique l'existence en son sein de structures qui vont permettre le passage de matériaux entre « l'extérieur » et « l'intérieur ».

LES 4 TYPES DE TRANSPORTS MEMBRANAIRES

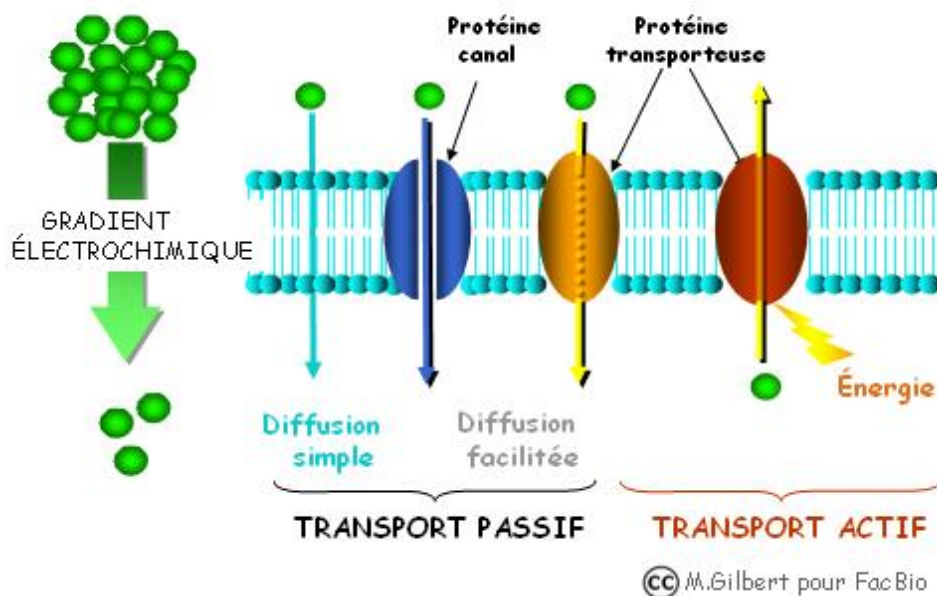


Figure II.10 : Les 4 types de transports membranaires : les 3 types de transports passifs, et le transport actif.

Les processus qui permettent l'absorption doivent être sélectifs, pour éviter qu'une cellule n'absorbe quelque chose qui lui est nocif ou qu'elle ne sait pas utiliser. Du fait de la très grande variété de composés qui peuvent être absorbés ou excrétés par les cellules, il existe des systèmes de transport variés.

Que ce soit entre une cellule et son environnement, ou entre 2 compartiments au sein d'une même cellule, il existe deux grands types de transport, les transports passifs et les transports actifs (Figure II.10).

Le transport peut concerner plusieurs espèces en même temps, on parle de cotransport*. Les espèces transportées peuvent alors être transportées dans le même sens (symport*) ou en sens inverse l'une de l'autre (antiport*). Certains transporteurs sont extrêmement spécifiques d'une espèce chimique, alors que d'autres sont plus polyvalents.

II.4.3.2.1. TRANSPORT PASSIF

Le transport passif se fait par la simple existence d'un gradient électrochimique, c'est-à-dire que le composé transporté est moins concentré dans un compartiment que dans l'autre et il se déplace alors naturellement, du compartiment où il est le plus concentré vers celui où il est le moins concentré. La vitesse de diffusion d'un composé est proportionnelle à la différence de concentration de ce composé entre les deux compartiments.

Ce type de transport peut se faire de deux manières, directement à travers la membrane cellulaire, par diffusion simple, qui concerne relativement peu d'espèces chimiques (souvent de petites molécules comme le dioxygène O_2 , ou le dioxyde de carbone CO_2) ou en faisant intervenir une protéine qui va faciliter le passage du composé échangé, on parlera alors de diffusion facilitée. Cette dernière peut se faire également de deux façons, soit à l'aide d'une protéine canal, une porine*, soit par une protéine de transport, une perméase*. L'eau ne peut traverser directement la membrane cellulaire qui est constituée de lipides et ses flux se font par une porine spécialisée, l'aquaporine*.

Remarque : lorsque c'est l'eau qui est considérée dans un processus de diffusion, on parle d'osmose. Cette notion sera importante lorsque nous aborderons la notion de pression osmotique, dans la description de l'environnement des cellules.*

II.4.3.2.2. TRANSPORT ACTIF

Le transport actif nécessite, quant à lui une protéine de transport et un apport d'énergie. Il peut aller à l'encontre d'un gradient électrochimique. L'énergie apportée peut être d'origine électrochimique, électrique, mécanique, lumineuse ou chimique.

Le transport actif est très important pour les cellules, car il permet de déplacer des solutés contre leur gradient de concentration. Ceci sera particulièrement important, par exemple, pour maintenir un gradient de concentration qui génèrera un flux à travers la membrane, qui pourra à son tour être utilisé comme source d'énergie mise à profit, par exemple, dans un autre processus de transport ou dans un processus de synthèse à travers une membrane, ou pour générer de l'énergie mécanique.

* Voir glossaire, page 199

II.5. ENVIRONNEMENT

Nous avons évoqué les échanges qu'entretiennent les cellules avec leur environnement, pour l'instant nous n'avons parlé que de ce qui se passe au sein des cellules, en le décrivant contraint par l'environnement, par rapport à l'énergie disponible sous la forme de la température, ou par rapport aux ressources en présence. Abordons maintenant les caractéristiques du milieu importantes pour la vie.

Si les organismes unicellulaires sont tributaires du milieu dans lequel ils se trouvent, les cellules des organismes pluricellulaires bénéficient d'un milieu plus stable, surtout chez les animaux, et ce sera particulièrement vrai avec les oiseaux et les mammifères, qui leur offrent en plus une température constante.

Les êtres vivants ont besoin d'énergie pour assurer leur survie. On distingue deux sources d'énergie, l'énergie lumineuse et l'énergie chimique. La chimie organique est la chimie du carbone. Tous les êtres vivants ont besoin pour se développer d'une source de carbone. Celle-ci peut être minérale (CO_2 ou ions dérivés) ou organique (sucres, lipides...). Pour synthétiser les protéines et les bases azotées de l'ADN et de l'ARN, il est également nécessaire de disposer d'une source d'azote. La vie sur Terre s'est développée en présence abondante d'eau, et tout organisme en requiert en grande quantité.

À partir de là, on trouve des microorganismes unicellulaires partout sur Terre, la diversité des environnements a produit, par le jeu de la sélection naturelle, des multitudes d'organismes adaptés à différents environnements, des glaces arctiques aux sources chaudes, des profondeurs de la terre aux dorsales océaniques. La variété des environnements s'accompagne d'une très large palette de caractéristiques. Ainsi, la salinité d'un milieu aura une forte influence pour la pression osmotique que subira un organisme. Lorsque la salinité à l'extérieur est la même que la salinité à l'intérieur de l'organisme, on parle d'un milieu isotonique*. Un milieu très riche en sel, qui exerce une forte pression osmotique, sera décrit comme hypertonique* et provoquera une importante sortie d'eau de la cellule, c'est ce que l'on appelle la plasmolyse. La déshydratation qui l'accompagne peut endommager gravement les constituants cellulaires. Par opposition, un milieu très peu salé sera décrit comme hypotonique et une cellule dans un tel milieu subira une très forte entrée d'eau, la cellule gonflée est qualifiée de turgescente, et si le phénomène est très important, il peut conduire à l'éclatement de la cellule.

Les cellules disposent de plusieurs artifices que nous avons déjà évoqués pour lutter contre la tonicité du milieu, la mise en place de vacuoles pour stocker de l'eau ou des solutés ou la formation des grains d'inclusion.

Une autre caractéristique de l'environnement, parmi les plus importantes, est la température du milieu, elle affecte particulièrement les organismes, mais aussi tous les êtres vivants dont la température interne varie avec celle du milieu extérieur. L'impact de la température sur le développement d'un organisme est surtout lié à la sensibilité thermique des réactions de son métabolisme. Nous avons vu que chaque enzyme a une température optimale de

* Voir glossaire, page 199

fonctionnement et à température trop basse elle cesse de fonctionner. Lorsque la température augmente, c'est la dénaturation de l'enzyme qui va généralement interrompre son fonctionnement. C'est donc tout le métabolisme des organismes qui est conditionné par la température du milieu. Les organismes homéothermes ont trouvé un moyen de se rendre indépendants de la température extérieure, en maintenant leur organisme à une température constante.

De plus, la membrane cellulaire étant constituée de lipides, elle se fige à température trop basse, ou fond et se déstabilise à température trop élevée. Dans les deux cas, les échanges de la cellule avec son environnement sont perturbés, ce qui affecte le développement de l'organisme. Si les températures optimales restent entre 15 et 40 °C pour une grande partie des organismes, on trouve sur Terre des microorganismes capables de se multiplier à la température extrême de -15°C (*Planococcus halocryophilus*) et à plus de 120°C (*Geogemma barosii*).

Trois derniers paramètres sont importants pour le développement des organismes. Le pH de l'environnement d'une part, l'atmosphère d'autre part et la pression pour finir.

Pour la plupart des organismes le pH idéal est un pH neutre. Par atmosphère, nous considérons les gaz, dans l'air ou dissous. Pour la plupart des organismes vivant à la surface de la terre, l'atmosphère normale avec sa pression et sa composition est l'atmosphère optimale, mais beaucoup de microorganismes ont des exigences particulières. Il en est de même pour la pression, pour la plupart des organismes la pression normale est la plus favorable, quelques organismes supportent des pressions plus fortes avec un record pour des microorganismes isolés dans la fosse des Mariannes qui ne peuvent se développer à une pression inférieure à 500 atmosphères (genres *Moritella*, *Shewanella* et *Colwellia*).

L'environnement naturel est soumis à une variété de rayonnements ionisants qui peuvent altérer les structures chimiques du vivant. La plus courante est le rayonnement ultraviolet. Le vivant s'est développé sur Terre et s'est adapté à un taux de radiation « normal », par exemple en développant des pigments pour se protéger des UV, et s'il existe peu d'organismes radiotolérants (avec un record *Deinococcus radiodurans*), il semble ne pas exister d'organisme radiophile. Les radiations, comme tous les stress qui peuvent être liés à l'environnement, ont vraisemblablement un rôle moteur dans l'adaptation et l'évolution que nous allons maintenant aborder.

II.6. NOTIONS D'ADAPTATION ET D'EVOLUTION

Adaptation et évolution sont deux phénomènes liés, c'est sur ces capacités d'adaptation que s'exerce la pression de la sélection naturelle, responsable de l'évolution. Ils s'observent à des échelles différentes, l'adaptation étant une réponse rapide à un stress quelconque, alors que l'évolution est une modification progressive des individus d'une population pour exploiter au mieux son environnement

II.6.1. ADAPTATION

L'adaptation consiste en la capacité à apporter une réponse adaptée à un changement rapide dans l'environnement. Cette réponse est rapide et réversible. Par exemple, quand il fait chaud on transpire, quand il fait froid, on tremble pour réguler notre température corporelle.

L'adaptation agit sur l'individu à l'échelle de quelques minutes ou de quelques heures, la mise en place d'un processus adaptatif est très brève comparée à la durée de vie de l'individu. Cette réponse est réversible puisqu'il s'agit d'un comportement adopté par l'individu et adapté à une situation donnée, qui est arrêté quand la situation change. En reprenant l'exemple de la réponse à la température extérieure, on tremble lorsqu'on a froid, on arrête lorsque la température remonte.

L'individu dont la capacité à se réchauffer ou à se protéger du froid est la plus efficace aura toutes les chances de mieux se porter, de mieux vivre ou d'avoir une plus grande descendance si le climat se dirige vers un nouvel âge glaciaire. Cette capacité à se protéger du froid pourra se faire de diverses façons. Par exemple, on peut imaginer la mise en place d'une couche protectrice, comme une fourrure plus dense, ou l'élaboration d'une couche de graisse isolante, ou encore l'augmentation de la thermogénèse au niveau des muscles ou au niveau des tissus adipeux. Les différents scénarii seront possibles, en fonction de ce que le patrimoine génétique de l'individu autorise. Par exemple, augmenter la thermogénèse nécessite d'avoir la possibilité d'augmenter le catabolisme, soit en augmentant la production d'enzymes impliquées au sein des cellules, soit en augmentant le nombre de cellules responsables.

Les capacités d'adaptation trouvent ainsi leurs fondements dans la génétique de l'individu. C'est en effet au niveau du patrimoine génétique que seront portées les informations qui vont permettre à un individu d'apporter une réponse adaptée à un changement de l'environnement, soit que le changement auquel on s'adapte a directement une influence sur l'expression des gènes, soit qu'il est lié à un processus de régulation.

Par exemple, une cellule exprime les gènes nécessaires à l'utilisation du lactose, si elle en rencontre dans son milieu, et éteint ces gènes, si ce sucre n'est plus présent. La capacité à utiliser un sucre sera définie par la capacité à absorber ce sucre, et donc à produire des transporteurs qui vont permettre l'incorporation de ce sucre. Une fois le sucre absorbé, il faudra que la cellule soit en mesure de l'utiliser, ce qui signifie avoir la machinerie enzymatique nécessaire à l'exploitation de ce sucre et donc avoir des gènes qui codent pour les transporteurs et les enzymes nécessaires. L'efficacité des transporteurs et des enzymes est liée à leur structure, elle-même définie par la séquence génétique. De même, la régulation des gènes impliqués sera liée à la structure et à la séquence génétique. Ainsi, la réponse sera plus ou moins efficace selon les individus et les mutations qu'ils peuvent porter.

II.6.2. EVOLUTION

L'évolution, définie par Charles Darwin en 1859, consiste en un changement graduel et irréversible des organismes, entre les générations (Figure II.11), en réponse à des

modifications progressives de l'environnement. L'évolution procède par la sélection naturelle qui favorise le développement des individus les mieux adaptés à un environnement donné ou dotés des meilleures capacités d'adaptation. On parle de pression de sélection.

Darwin a eu l'intuition que la variabilité qui existe entre les individus était la clé de l'adaptation et que les contraintes environnementales sélectionnaient les individus les plus aptes. Les jeunes les plus fragiles d'une espèce étant rapidement éliminés pour laisser place aux plus forts, qui ont le plus de chance d'assurer une descendance nombreuse. Il a également introduit le concept de sélection sexuelle en observant que chez de nombreuses espèces un mâle dominant s'octroie un « harem » et lui seul se reproduira tant, qu'il conserve sa place de mâle dominant. Il est alors le seul à avoir une descendance. Chez d'autres espèces, les femelles choisissent leur partenaire en fonction de critères de beauté ou de force spécifiques. La désignation du mâle dominant ou le choix du mâle par une femelle suscite une grande compétition entre les mâles.

Il manquait à Darwin la compréhension de l'hérédité. En ce sens, la conception de Lamarck, qui décrit une modification chez un individu comme résultant d'une volonté et transmissible par la reproduction sexuée était une piste. Il lui a manqué la connaissance des travaux de Mendel, qui ont posé les bases de la génétique, pour comprendre l'origine de ces capacités d'adaptation, et comme origine des variations qui vont permettre la sélection sexuelle.

Les concepts de la génétique ont été intégrés à la notion d'évolution pour construire la théorie synthétique de l'évolution. Cette théorie moderne s'enrichit des nouveaux savoirs, à mesure qu'ils se font jour. Ainsi, l'importance des migrations et du cloisonnement géographique sont apparus comme autant de nouveaux moteurs de l'évolution. La diversité génétique de l'échantillon qui migre n'est pas forcément représentative de la population totale, et le processus de migration peut favoriser certains variants, si bien que, dans le nouveau milieu, le pool génétique sera réduit, apportant de nouvelles contraintes à la sélection naturelle. L'isolement géographique limite le brassage génétique et donc l'apparition de nouveaux variants et augmente également les pressions de la sélection naturelle. Et là où Lamarck pensait qu'il y avait une volonté dans l'adaptation et l'évolution, on sait désormais que la sélection naturelle est aveugle et est fortement liée au hasard. En effet, un jeune peut être porteur d'un patrimoine exceptionnel qui lui assurerait une grande descendance, mais encore faut-il qu'il atteigne l'âge de la reproduction. Et ce ne sera pas le cas si un manque de soins maternels l'empêche de se développer pleinement, ou s'il est victime de prédation, ou encore si par malchance une catastrophe naturelle, comme un incendie, frappe l'endroit où il vit.

L'apparition de ces variants est liée, en vertu de la génétique, à des mutations, qui vont améliorer des fonctions, faire perdre des capacités, ou faire naître de nouvelles facultés. Ces mutations peuvent être liées à des erreurs lors de la copie du génome de l'individu, ou être induites par des facteurs externes, comme les rayonnements ionisants produits par la radioactivité naturelle, les rayons cosmiques, ou encore les rayons ultraviolets.

Une niche écologique laissée vacante pourra ainsi devenir le foyer d'abord d'un variant, puis, après un certain nombre de générations, le variant deviendra espèce.

L'évolution permet ainsi à une espèce d'émerger dans un milieu donné en produisant des individus toujours plus adaptés à leur environnement. Les espèces se modifient peu à peu au fil des générations et des variations qui existent entre les individus, sous la contrainte des changements dans l'environnement et de la compétition entre les individus. Ce phénomène est très lent, se déroulant entre les générations, il n'apparaît que sur des durées très longues comparées à la vie de l'individu.

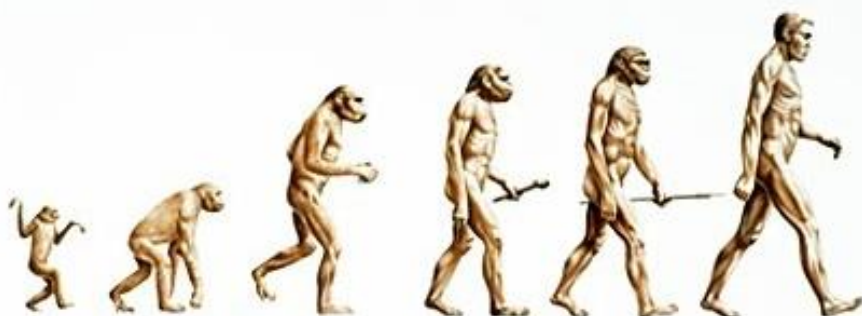


Figure II.11 : "The march of progress" (Howell, *Early Man*, 1965).

L'évolution n'est plus considérée comme un processus continu mais comme une succession de phases où les populations sont à l'équilibre, entrecoupées de périodes d'intense spéciation. On parle d'équilibres ponctués introduits en 1972 par Stephen Jay Gould et Niles Eldredge.

L'évolution consiste en une marche en avant irréversible. Même si certains cas (par exemple une fonction perdue qui réapparaît dans l'arbre évolutif) pourraient être perçus comme un retour en arrière, la façon dont s'opère le second changement, la réapparition de la fonction, est différente du premier, la perte de la fonction.

Au sein de lignées d'organismes différents, des processus évolutifs peuvent, à travers des stratégies différentes, donner des fonctions aux résultats similaires, on parle de convergence évolutive. Le vol en est un bon exemple, il est différent chez les insectes, les oiseaux ou les mammifères.

II.7. CADRE EXPERIMENTAL

Avant de clore ce chapitre, portons notre attention sur deux éléments qui nous seront utiles par la suite, la culture cellulaire d'une part et la radiobiologie d'autre part. Le premier parce que la culture cellulaire est un élément facile à réaliser en laboratoire, qui traduit déjà l'ensemble des processus cellulaires à l'échelle d'une population. Cette technique de laboratoire nous servira de mètre-étalon pour valider le comportement du modèle que nous proposerons et pour le paramétrer et il sera facile de soumettre des cultures cellulaires à l'irradiation, pour en observer l'effet. Ce qui amène le second élément, la radiobiologie. Cette notion exprimée dans la problématique nécessite quelques compléments d'information, à la suite de la description des systèmes biologiques qui vient d'être faite, pour préciser l'effet

des rayonnements ionisants sur le matériel biologique et les conséquences qui peuvent être observées.

II.7.1. ENVIRONNEMENTS EXPERIMENTAUX, NOTION DE CULTURE CELLULAIRE

Nous avons déjà abordé sommairement la description d'un environnement considéré favorable à la croissance cellulaire, apportons quelques précisions.

Certains microorganismes ont besoin de concentrations en sels minéraux élevées pour se développer de façon optimale, on parle d'organismes halophiles*. Des organismes qui parviennent à se développer avec des concentrations en sels importantes, mais de façon moins efficace que dans un milieu moins salé, seront dits halotolérants*.

Pour beaucoup de microorganismes, la croissance optimale sera obtenue dans des conditions de pH comprises entre 5,5 et 8. On décrira ces organismes comme neutrophiles*, et cette gamme de pH correspond à ce que l'on appelle en général le pH physiologique. D'autres organismes auront une préférence pour un pH acide, 0 à 5,5, on parlera alors d'organismes acidophiles*. A l'inverse, des organismes dits alcalophiles* auront leur pH optimum de croissance à une valeur comprise entre 8 et 11.

La température a une forte influence sur la croissance cellulaire, les températures extrêmes qui autorisent la croissance pour un microorganisme donné sont les températures cardinales. Entre les deux, il existe une température optimale de croissance. Ces températures cardinales n'ont pas une valeur fixe, mais dépendent d'autres paramètres de l'environnement. Les températures cardinales et optimales varient de plus considérablement suivant les microorganismes, les optimums s'échelonnant généralement de 0°C à 75°C.

Les microorganismes ayant une température de développement optimale vers 10-15°C environ sont dits psychrophiles*. Ces organismes commencent à perdre leurs fonctions vers 20°C. Ces microorganismes sont trouvés en abondance dans les océans. Des organismes qui peuvent se développer à des températures inférieures à 7°C, mais avec une température optimale variant entre 20 et 30°C, sont dit psychrophiles facultatifs, cela concerne beaucoup de champignons et bactéries du sol, par exemple.

Les mésophiles* sont les microorganismes dont la température optimale de croissance se situe vers 35-40°C, avec des températures minimales et maximales de l'ordre respectivement de 20°C et 45°C. Les microorganismes appartenant à la flore normale des animaux appartiennent à cette catégorie, de même que la plupart des pathogènes humains.

Lorsque la température minimum de développement d'un microorganisme est d'environ 45°C, avec un optimum compris entre 55 et 65°C on parlera d'organismes thermophiles*. On les rencontre dans les sources chaudes, le compost... Les hyperthermophiles* sont capables de se développer jusqu'à des températures supérieures à 120°C, mais supportent mal des températures inférieures à 55°C, avec une température optimale comprise entre 80 et 113°C. On les rencontre dans les sources chaudes, essentiellement sur les dorsales océaniques.

* Voir glossaire, page 199

Ces notions ont une importance pour envisager l'influence de l'environnement sur la croissance que l'on peut obtenir lors d'une expérience en laboratoire. Pour une lignée de cellules donnée, il faudra étudier la croissance dans les conditions choisies pour évaluer sa bonne modélisation. La mise en œuvre de la culture cellulaire peut se faire sur un milieu solide ou en milieu liquide. Les éléments qui peuvent être étudiés sont alors différents, la culture en milieu solide ne permet pas l'évaluation des caractéristiques dynamiques de la croissance, mais autorise une étude d'aspects physiques (forme, adhésion, ...) et physiologiques (pigmentation, mobilité, ...) des cellules, alors que la culture en milieu liquide ne permet pas l'observation de ces derniers aspects, mais se révèle toute indiquée pour étudier la cinétique de la croissance.

II.7.1.1. CULTURE EN MILIEU SOLIDE

La culture en milieu solide sert à isoler les microorganismes présents dans un mélange, afin de pouvoir étudier leurs caractéristiques individuellement et de pouvoir les identifier. Lorsqu'on réalise une culture en milieu solide, la technique d'isolement sert à séparer les cellules de l'inoculum et on considère ainsi que toutes les cellules d'une colonie sont issues de la même cellule.

Les colonies des microorganismes présentent des caractéristiques qui sont des critères de classification et d'identification morphologiques (Figure II.12). Citons par exemple :

- la forme de la colonie ;
- le relief de la colonie ;
- les bords de la colonie ;
- la présence de pigments.

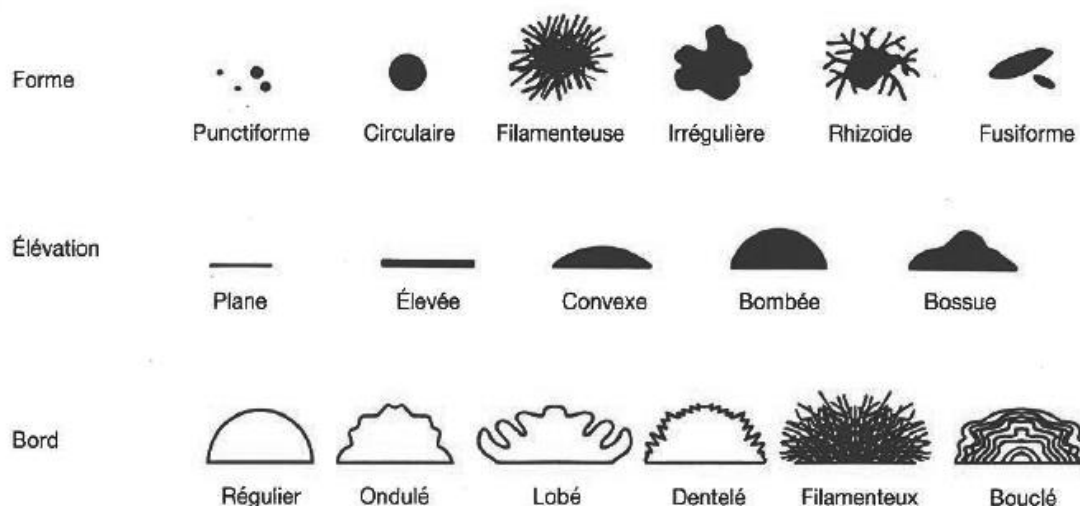


Figure II.12 : Les critères morphologiques d'identification des microorganismes en culture sur milieu solide (Prescott et al., 2010 [2]).

Ces aspects sont caractéristiques de l'espèce et servent de critères d'identification des microorganismes. Ils sont donc un aspect du phénotype bactérien, et en tant que tel, ils représentent l'expression de certaines propriétés génétiques.

Une colonie de forme différente, d'une autre taille, d'une couleur anormale, etc... est signe d'une bactérie différente. Si l'inoculum était pur, avec une seule espèce de bactéries, ce phénotype différent est synonyme d'un changement survenu dans la cellule à l'origine de la colonie, et donc vraisemblablement d'une mutation.

II.7.1.2. CULTURE EN MILIEU LIQUIDE

La culture en milieu liquide de cellule peut se faire en milieu fermé, c'est ce que l'on appelle culture en batch* ou discontinue, et en milieu renouvelé, c'est ce que l'on nomme culture en bioréacteur* ou chémostat* ou encore culture continue. Un milieu de culture liquide est aussi appelé bouillon de culture et la croissance cellulaire se traduira par le développement d'un trouble dans le milieu.

Dans la culture en batch, un volume de milieu de culture stérile est introduit et inoculé avec la lignée de cellules dont on souhaite étudier la croissance. La croissance ralentira et s'arrêtera d'elle-même lorsque les nutriments viendront à manquer ou que les déchets commenceront à trop s'accumuler dans le milieu. Lors d'une culture en bioréacteur, comme on apporte du milieu neuf, stérile, en permanence, la croissance peut être maintenue très longtemps. L'apport de milieu frais est compensé par la ponction continue de milieu. Ceci assure un maintien de la culture dans un milieu aux caractéristiques physico-chimiques constantes.

La croissance cellulaire peut être suivie par plusieurs méthodes, le dénombrement de toutes les cellules, le dénombrement des cellules viables, le suivi de la masse cellulaire totale, etc... Pour les procaryotes, une technique souvent utilisée est la spectrophotométrie, qui permet de mesurer la densité de cellules dans un milieu liquide, on parlera de densité optique* ou DO. Cette dernière est une grandeur sans unité que nous utiliserons ultérieurement. L'évaluation de la densité d'une population cellulaire par mesure de la DO ne fonctionne que lorsqu'on utilise un milieu liquide agité. Couramment, on estime qu'une valeur de 1 de DO correspond à une densité de 10^7 à 10^9 cellules par millilitre.

Le cycle cellulaire nécessaire pour qu'une cellule mère se développe et donne deux cellules filles se traduit dans la croissance par un temps mis par une population pour doubler de taille. Ce temps est appelé temps de génération G. Il est relié au taux de croissance μ par la relation $\mu = \frac{1}{G}$. Dans les conditions optimales, le cycle de croissance est plus court que dans toute autre configuration. C'est dans ces conditions que l'on peut observer la valeur maximale de μ (ou la valeur minimale de G) pour un organisme donné.

La croissance des microorganismes a été caractérisée avec des cultures en batch. Un volume de milieu stérile est inoculé avec la lignée de cellules étudiée. Elle s'articule alors en quatre phases décrites en 1909 par Janet Lane-Claypon, la phase de latence (Figure II.13-1), la phase exponentielle (Figure II.13-2), la phase stationnaire (Figure II.13-3) et la phase de déclin (Figure II.13-4).

* Voir glossaire, page 199

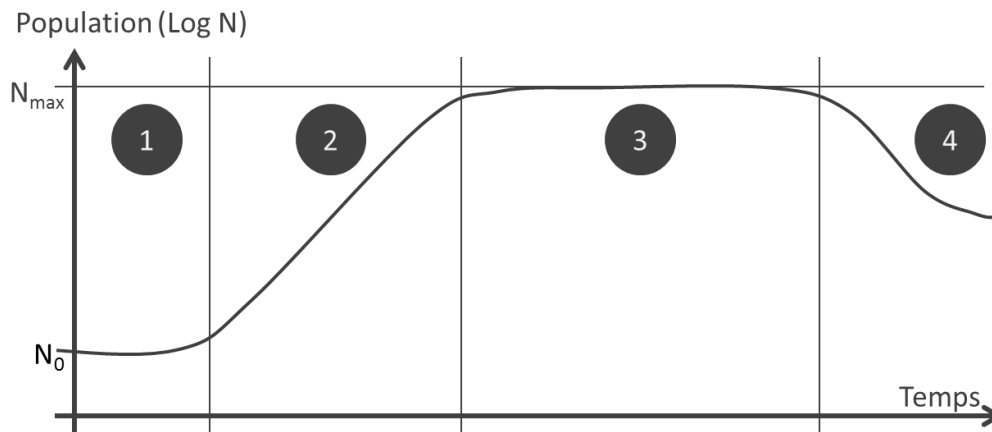


Figure II.13 : Les principales phases de la croissance cellulaire observées lors d'une culture en batch, inspiré de Buchanan (Buchanan, 1918 [4]).

Les caractéristiques de la cinétique de croissance dans un environnement donné constituent ce qui peut s'apparenter à un phénotype. Tout changement dans les paramètres de la croissance sera signe d'une modification des capacités des cellules sollicitées dans les conditions expérimentales et sera signe d'une altération des fonctions cellulaires, et donc probablement d'une mutation.

A. PHASE DE LATENCE

Cette phase a une durée variable, elle peut être très longue ou très courte, selon les conditions et selon le type de cellule. Elle peut même ne pas être visible dans certaines conditions. Cette phase peut avoir plusieurs origines. Si la population de cellules qui est utilisée pour l'inoculation est constituée de cellules âgées et abimées, dans un mauvais état métabolique, il leur faudra un temps pour se remettre dans un état métabolique satisfaisant, afin de pouvoir commencer à proliférer. Une autre cause fréquente est liée à l'adaptation des cellules à leur nouveau milieu. Cela s'observe lorsque les cellules qui servent à l'inoculation sont prélevées dans un milieu différent de celui qui va servir à réaliser la culture. Lors de cette phase, en général, le nombre de cellules reste constant, mais la biomasse augmente.

Si les cellules utilisées pour l'inoculum sont ponctionnées dans une culture qui est dans la phase exponentielle de croissance et dans un milieu identique à celui qui va accueillir la nouvelle culture, que les conditions physico-chimiques restent les mêmes, la phase de latence sera réduite, voire absente.

Au cours de cette phase, la taille de la population reste pratiquement constante, le nombre de cellules présentes reste environ égal au nombre de cellules présentes dans l'inoculum N_0 .

B. PHASE EXPONENTIELLE

Au cours de cette phase, la vitesse de division cellulaire est constante et au plus haut niveau permis dans les conditions de culture. Toutes les cellules sont dans un état métabolique similaire.

C'est au cours de cette phase que la mesure du temps de génération peut être réalisée, le taux d'accroissement de la population étant pratiquement constant. Insistons sur le fait que la cinétique de croissance observée est dépendante de l'espèce cultivée et des conditions de la culture, les paramètres de la croissance pourront être très différents dans des conditions différentes.

Au cours de cette phase, le temps de génération est le plus court observable pour la cellule cultivée dans les conditions expérimentales utilisées. Il est relié au taux de croissance maximum μ_{\max} .

C. PHASE STATIONNAIRE

L'entrée dans cette phase dépend probablement de plusieurs facteurs mal connus. La densité de la population semble jouer un rôle, mais de façon plus évidente elle survient lorsque le milieu devient moins favorable, c'est-à-dire que l'accumulation des déchets atteint un niveau où elle perturbe la croissance des cellules ou lorsqu'un ou plusieurs nutriments viennent à manquer.

Au cours de cette phase le nombre de cellules viables a sa valeur maximale N_{\max} , et est pratiquement constant. La croissance n'est pas pour autant arrêtée, c'est en réalité que la population atteint un équilibre entre le nombre de cellules qui meurent et le nombre de cellules qui apparaissent.

Lorsque la culture est dans cette phase, les cellules régulent plus finement leur métabolisme et renforcent leur structure, de sorte qu'elles résistent mieux à une carence et à des stress physico-chimiques (stress oxydatif, stress osmotique, ...).

D. PHASE DE DECLIN

La phase de déclin, ou phase de mortalité, se caractérise par le fait qu'il meurt plus de cellules que n'en naissent de nouvelles. Cela peut être dû à un milieu qui devient nocif, par l'accumulation d'une trop grande quantité de déchets, ou à la carence en nutriments.

Chez les bactéries, cette phase peut se prolonger longtemps, c'est ce que l'on appelle phase de déclin prolongé, lorsque le recyclage de la matière organique des cellules mortes est suffisamment efficace pour permettre un maintien d'une légère prolifération cellulaire. Cette phase de déclin prolongé s'apparente alors à une phase stationnaire, mais avec une population beaucoup moins dense. Par un processus de sélection naturelle, la population qui se maintient devient de moins en moins exigeante, de plus en plus efficace pour recycler les déchets accumulés dans l'environnement et de plus en plus tolérante vis-à-vis de la nocivité du milieu.

II.7.2. RADIOBIOLOGIE

La radiobiologie est une discipline qui s'intéresse aux effets de la radioactivité sur le matériel biologique. Les rayonnements ionisants peuvent avoir des effets à différents niveaux. Ils occasionnent des dommages directs, en interagissant avec la matière des constituants

cellulaires. Ces dommages directs génèrent des espèces chimiques très réactives, les radicaux libres, qui sont de puissants oxydants. Ces derniers vont réagir rapidement avec d'autres molécules au sein d'une cellule, occasionnant de nouveaux dommages, dits dommages indirects.

L'irradiation peut être d'origine naturelle, liée à la radioactivité présente dans l'environnement, accidentelle, liée à des accidents, comme Tchernobyl en 1989 ou Fukushima en 2011, ou voulue, comme ce sera le cas dans le domaine médical, avec les radiothérapies. Dans la nature, la radiation ionisante la plus fréquente est le rayonnement ultraviolet. La lumière ultraviolette la plus délétère est à 260nm de longueur d'onde, celle-ci étant la plus absorbée par l'ADN. Même si les microorganismes résistent mieux à l'irradiation que les organismes supérieurs, il n'existe que peu d'organismes qui soient réellement radiotolérants.

Quelle que soit l'origine de l'irradiation, tous les constituants cellulaires peuvent être touchés, tant par les dégâts directs que par les dégâts indirects. L'impact sur la vie de la cellule dépendra de l'importance des dommages subis. Dans le cadre d'une radiothérapie, l'objectif sera de tuer les cellules cancéreuses, en limitant les dégâts sur les tissus sains environnants.

À l'échelle d'une cellule, si ce sont les effecteurs que sont les protéines qui sont endommagés, la cellule a de grandes chances de s'en remettre, du moins tant que sa machinerie reste en suffisamment bon état pour réparer ou resynthétiser les protéines dégradées, la cellule pourra alors se rétablir plus ou moins rapidement. Si trop d'effecteurs sont détruits lors de l'irradiation et qu'il n'en reste plus suffisamment pour assurer une maintenance et une restauration des constituants cellulaires, les effets seront létaux.

Si les rayonnements ionisants ont touché le patrimoine génétique, le devenir de la cellule dépendra de l'importance des dégâts. Si les dommages sont peu importants ou qu'ils sont localisés dans des zones de faible importance du génome, il y a de fortes chances que la cellule puisse les réparer. Si les dommages sont plus importants, ou touchent des zones importantes du génome, la réparation sera peut-être possible, mais s'accompagnera peut-être d'erreurs qui se traduiront par des mutations et donc, potentiellement, par des modifications ou des pertes de fonctions cellulaires. Si les lésions sont très importantes ou affectent des séquences capitales du génome, il est très probable que la cellule n'y survivra pas.

Les capacités à résister ou à se remettre des lésions dues aux radiations sont très variables d'une espèce à l'autre. Elles seront liées aux capacités régénératives dont disposent les cellules, mais aussi à la structure biochimique de leurs constituants. *Deinococcus radiodurans* peut résister à 5 000 fois la dose de rayonnements qui serait considérée comme létale pour un être humain, et alors qu'on estime que deux à trois cassures double-brin du chromosome d'*E. coli* suffisent à provoquer sa mort ou des mutations importantes, *D. radiodurans* est capable de survivre à une centaine de cassures double-brin de ses chromosomes.

Au niveau du génome, il est difficile de prédire si un gène sera touché, d'autant plus que le nombre de gènes est très variable d'une espèce à l'autre, et que la variabilité interindividuelle joue sur la longueur des gènes, leur position exacte (début d'un gène à plus ou moins une

centaine de bases près). L'observation du phénotype pourra ne pas être suffisante pour révéler une mutation. Comme ce dernier résulte de l'interaction entre le génotype et l'environnement, les conditions d'observation peuvent avoir une influence, et la difficulté sera encore plus grande chez les organismes qui disposent de plusieurs copies d'un même gène. Si l'une est affectée, les autres continueront à fonctionner et aucun effet ne sera visible.

Évaluer l'énergie de la radiation impliquée dans l'interaction avec le matériel biologique et donc la quantité de dommages qui peuvent être subis est complexe, particulièrement dans le cadre d'une radiothérapie, si la zone ciblée est en profondeur dans les tissus. Il sera d'autant plus difficile de savoir quelle énergie atteindra réellement les cellules tumorales ciblées qu'elle dépendra de l'absorption des radiations par les tissus environnants, et elle-même dépendra de la physiologie du patient.

Affiner la connaissance des mécanismes de radiobiologie et savoir les modéliser permettra de faire des tests *in silico*, qui faciliteront la mise en œuvre de traitements.

Dans les expériences de radiobiologie menées *in vitro* sur des cultures cellulaires, tout changement dans les caractéristiques de la croissance sera signe d'une mutation. Si l'on s'en tient à ce cadre expérimental et que l'on se cantonne à l'observation du comportement en culture de cellules irradiées, nous pouvons envisager trois cas de figure possibles : soit la cellule continue à vivre et à se multiplier, comme si de rien n'était (pas de perturbation visible), soit la cellule continue à vivre et à se multiplier mais à un rythme différent de celui attendu (le temps de génération est modifié), soit la cellule ne se multiplie plus.

Dans le premier cas de figure, lorsque la cellule continue à se multiplier au rythme attendu on peut considérer deux situations : soit l'irradiation n'a pas fait de dégâts sur une séquence importante du génome dans les conditions expérimentales, soit les dégâts sur le génome ou sur les protéines occasionnés par l'irradiation ont pu être réparés rapidement et sans difficultés. Le cycle de vie cellulaire pourrait toutefois être perturbé dans d'autres conditions de culture.

Dans le second cas de figure, nous pouvons envisager deux explications pour décrire le fait que la cellule se multiplie à un rythme différent de celui attendu. Même si c'est très peu probable, il se pourrait qu'on observe un rythme de croissance accéléré. Dans ce cas, l'irradiation a induit une mutation qui se révèle favorable dans les conditions de culture utilisées, alors que d'autres conditions ne l'auraient peut-être pas révélée.

Si la cellule se multiplie plus lentement que ce à quoi on s'attendrait, on pourra déduire que l'irradiation a perturbé un ou plusieurs gènes impliqués dans les fonctions cellulaires sollicitées dans les conditions de la culture, que ce soit la capacité d'absorption d'une ressource ou sa capacité d'exploitation. L'altération peut simplement avoir touché les protéines, et elle disparaîtra peu à peu, ou alors ce sont les gènes responsables qui ont muté. On peut aussi supposer que ce sont les gènes impliqués dans le processus de la division cellulaire qui ont été altérés.

Dans le troisième cas de figure envisagé, la cellule ne se multiplie plus. Soit la cellule a été tuée, soit elle a perdu sa capacité à se diviser. Elle mourra alors effectivement en vieillissant, mais jusque-là elle peut vivre normalement. Il se peut aussi que la cellule meure

prématurément, suite à la perte d'une ou plusieurs fonctions cellulaires et on peut alors proposer quatre hypothèses pour l'expliquer.

La première est que la cellule a perdu la capacité à dupliquer ses constituants pour se préparer à la division. Dans la seconde hypothèse, c'est la possibilité de croissance de la cellule qui est affectée, et si la cellule n'atteint pas la taille critique, le processus de division ne sera pas enclenché. En troisième lieu, il se pourrait que seul le processus de division ait été affecté, la cellule grandira et grossira, dupliquera son matériel pour préparer une phase de division, mais les copies resteront au sein d'une même cellule devenue géante. La cellule finira par mourir en éclatant, lorsque la pression de ses constituants dépassera la résistance de son enveloppe. Ces trois hypothèses sont indépendantes des caractéristiques de la culture, à la différence de la dernière hypothèse que l'on peut formuler : la cellule peut avoir perdu un processus biochimique indispensable dans les conditions utilisées et ne peut plus utiliser les ressources.

II.8. CONCLUSION

Les systèmes biologiques sont complexes par nature, tant par l'importance du nombre d'éléments en interaction, que par l'impossibilité d'expliquer un tel système en se limitant à la description de ses éléments constitutifs. Dans cette définition même, la notion de multi-échelle transparaît, un système biologique peut être, et même doit être, décrit en prenant en considération ses différents niveaux d'organisation.

Pourtant plusieurs caractéristiques peuvent être dégagées de la description qui vient d'être réalisée. Même si l'on ne considère que les organismes les plus simples, trois niveaux d'observation semblent particulièrement importants, à la lumière de l'objectif radiobiologique :

- le niveau moléculaire, qui va décrire tous les processus biochimiques à l'œuvre dans le métabolisme et au niveau duquel se joueront les interactions entre rayonnement et matériel biologique ;
- le niveau cellulaire où va pouvoir s'observer le vivant et sa capacité de résilience à un stress ;
- le niveau de la population où va pouvoir se traduire la sélection des éléments les plus aptes à s'adapter à un stress.

Focaliser notre attention sur les bactéries simplifie l'objectif de modélisation en négligeant certains aspects structurels des cellules eucaryotes et constitue ainsi un premier jalon raisonnable pour l'étude des effets des rayonnements sur le vivant. De plus, leur culture étant simple et rapide à mettre en œuvre en laboratoire, il est relativement facile d'obtenir d'expériences de culture cellulaire des données quantifiées, sur lesquelles se baser pour paramétrer les modèles que l'on peut étudier.

CHAPITRE III. ETAT DE L'ART : DES MODELES AUX PLATEFORMES

III.1. Introduction.....	65
III.2. Modèles de population cellulaire : différentes approches.....	67
III.2.1. Modélisation à population pour les processus biologiques.....	69
III.2.2. Modélisation individu-centrée pour les systèmes biologiques.....	73
III.2.2.1. Automates cellulaires.....	74
III.2.2.2. Modèle de Potts.....	77
III.2.2.3. Système Lattice-Gas.....	79
III.2.2.4. Systèmes multi-agents.....	81
III.2.3. Bilan sur les types de modèles.....	82
III.3. Modèles pour les processus biologiques.....	83
III.3.1. Processus biochimiques.....	86
A. SimCell.....	86
B. V-Cell.....	88
C. Snoopy.....	90
III.3.2. Adaptation.....	91
A. Bacsim.....	91
B. Indisim.....	93
III.3.3. Evolution.....	96
A. Tierra.....	97
B. Evolved virtual Creature.....	98
C. Avida/Avida-ED.....	99
D. AEVOL.....	100
III.3.4. Synthèse de ce tour d'horizon des modèles.....	102
III.4. Plateformes.....	103
III.4.1. Formalismes.....	103
A. SBML : System Biology Markup Language.....	103
B. CellML :.....	104
III.4.2. Plateformes dédiées.....	104
III.4.3. Plateformes génériques.....	105
A. E-Cell.....	105
B. CompuCell3D.....	106
C. Morphex.....	106
III.5. Conclusion.....	107

III.1. INTRODUCTION

A la lumière de notre problématique, la description du vivant nous aura permis d'identifier un certain nombre de caractéristiques qu'il nous faudra garder à l'esprit pour notre objectif de modélisation.

Nous nous centrons sur l'entité fonctionnelle du vivant, la cellule. Il nous faut pouvoir gérer plusieurs échelles autour, ce qui se passe aux échelles inférieures au sein même des cellules, ce qui se passe aux échelles supérieures, au niveau de massifs cellulaires, de tissus, de colonies, etc... En termes de temps, nous avons une palette de processus qui se déroulent sur des durées très différentes, de l'ordre de la microseconde, voire moins pour les processus biochimiques à des durées d'ordre géologique pour l'évolution, ou en simplifiant, à des durées qui sont très brèves ou extrêmement longues par rapport à l'existence d'une cellule ou d'un organisme.

En nous centrant sur l'échelle cellulaire, il nous faut pouvoir gérer chaque cellule individuellement, son patrimoine génétique, son fonctionnement, ses interactions avec son environnement, sa survie ou sa mort, ses déplacements, etc... que chacune puisse être autonome. Puisque chaque cellule interagit avec ce qui l'entoure, échange de la matière, de l'énergie et de l'information avec son environnement, il faut pouvoir gérer ce qu'il se passe dans ce milieu dans lequel elle est placée. Il faut ainsi prévoir la possibilité de gérer des flux entre les cellules, et des flux entre les cellules et leur environnement, et au sein de l'environnement.

On le voit en listant ces caractéristiques, pour répondre à notre objectif, il nous faut être en mesure de décrire des processus très rapides, à petite échelle spatiale, aussi bien que des phénomènes lents, qui se déroulent sur de vastes territoires, mais en même temps, des processus rapides, qui se produisent sur des espaces larges, ainsi que d'autres se déroulant lentement, mais sur des dimensions spatiales restreintes.

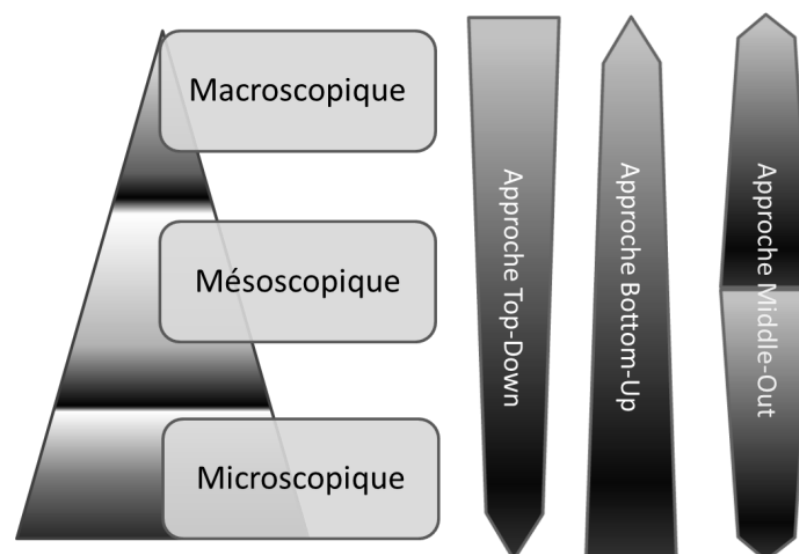


Figure III.1 : La pyramide des échelles et les approches pour l'exploration des systèmes complexes.

Deux approches peuvent alors se dégager spontanément, l'approche dite « top-down » et celle dite « bottom-up ». Il existe aussi une approche intermédiaire, dite « middle-out » ((Figure III.1).

Dans la première, on considère le système dans son ensemble et on va le découper en ses parties, et ce, de plus en plus finement jusqu'à obtenir des systèmes élémentaires. C'est l'approche que nous avons eu en décrivant un écosystème dans la description du système biologique, une forêt est constituée d'arbres, chaque arbre est constitué d'un tronc, de branches et de feuilles, etc... C'est du réductionnisme, on part d'un modèle global et par réductions successives on atteint le niveau de finesse voulu. Une ville est constituée de bâtiments, eux-mêmes constitués avec des murs, eux-mêmes construits avec des briques, etc... Dès lors que l'on connaît le modèle global on peut envisager de détailler ses composantes.

L'approche bottom-up se fait en sens inverse, c'est de l'intégration. Dans ce cas le travail, va consister à partir des échelles les plus petites, en termes d'espace et de temps, et à les additionner jusqu'à obtenir les échelles plus larges. Ceci nécessite de bien connaître tous les comportements microscopiques pris en compte. Le nombre d'éléments qui peuvent alors être ajoutés devient rapidement gigantesque, et les capacités de calculs requises pour modéliser le système de façon exhaustive par cette approche deviennent rapidement colossales. Par exemple, si l'on fait une rapide hypothèse, une cellule de $1 \mu\text{m}^3$, soit de 1 fL constituée d'eau pure aurait une masse de 1 pg. En vertu du nombre d'Avogadro, $6,02 \times 10^{23}$, l'eau ayant une masse molaire de 18 g.mol^{-1} , une telle cellule représenterait $3,34 \times 10^{10}$ molécules d'eau. Modéliser une seule cellule d'un tel type en partant des molécules d'eau nécessiterait donc de modéliser $3,34 \times 10^{10}$ particules. Dans cette approche, il faut avoir la connaissance de tous les éléments que l'on souhaite modéliser, et la connaissance de toutes les interactions pour construire le modèle le plus juste. Il faut savoir de quelle échelle on part et à quelle échelle on souhaite arriver : avec des briques, on fait des murs, avec des murs, on fait des bâtiments, avec des bâtiments on fait des villes...

Il existe une approche intermédiaire dite « middle-out ». Celle-ci consiste à partir d'une échelle quelconque, et à voir ce que l'on peut affiner comme éléments dans les échelles inférieures, et comment les éléments de l'échelle de départ s'agencent pour former le système. Plus simplement, cette approche combine une approche top-down et une approche bottom-up menées depuis l'échelle choisie comme point de départ. On parlera d'ailleurs d'une démarche « middle-up » pour intégrer les processus de l'échelle de départ vers les phénomènes observés aux niveaux supérieurs, et d'une démarche « middle-down » pour la réduction des connaissances de l'échelle de départ en processus de plus bas niveaux. Un bâtiment est construit avec des murs construits avec des briques. Des bâtiments ensembles, ça donne une ville.

En ayant décidé de nous focaliser sur la cellule comme entité élémentaire fonctionnelle du vivant, cette dernière approche semble naturelle : de quoi est faite une cellule, jusqu'à quel niveau de détails descendrons nous, et comment agrégerons nous les informations des échelles que nous négligerons car trop fines, et comment interagissent plusieurs cellules entre elles, comment se forment les comportements observés à l'échelle de la population.

La difficulté qui se pose alors est que le vivant est un système complexe. La définition d'un système complexe indique que le comportement de l'ensemble ne peut être prédit du comportement de ces constituants, il existe une valeur ajoutée qui n'est visible que lorsque les éléments constitutifs du système sont en interaction et qui est perdue lorsqu'on étudie les parties individuellement (Lesne, 2009 [5]). Beslon considère « que la biologie, qui a longtemps fait un usage quasi-exclusif de la méthode réductionniste avec les succès que l'on sait, arrive aujourd'hui à un tournant de son histoire. La masse de données accumulées sur les éléments des systèmes biologiques a indirectement ouvert une boîte de Pandore : il n'est plus possible de considérer que la compréhension de l'activité d'un système biologique proviendra de cette accumulation, pour la simple et bonne raison que l'organisation des constituants n'est pas contenue dans la description de ces mêmes constituants, aussi détaillée soit-elle. Or, les propriétés du système, du "tout", dépendent au moins autant de cette organisation que des caractéristiques locales de ses constituants » (Beslon, 2008 [6]).

Et effectivement, par exemple un être humain de 70 kg correspond à 43 kg d'oxygène, 7 kg d'hydrogène, 16 kg de carbone, 1,8 kg d'azote, 0,78 kg de phosphore, 1 kg de calcium, le reste se répartissant en soufre, sodium, chlore, potassium, magnésium, silice, fer, fluor, zinc, et quelques éléments à l'état de traces (Stumble Upon [7] via BioNumbers [8]). Pour autant réunir tous ces éléments dans une baignoire ne donnera pas un être humain. La nature, les capacités, les caractéristiques d'un organisme humain ne peuvent être observées si l'on étudie chaque élément séparément, quand bien même la chimie de ces espèces serait parfaitement connue. De même un simple organe ne peut être compris en observant seulement les cellules qui le composent individuellement, par exemple observer un neurone isolé ne rend pas compte du fonctionnement du cerveau. Et de la même manière, étudier le comportement d'un individu ne permet pas de présager du comportement d'une population entière.

Dès lors, comment définir les échelles inférieures au niveau cellulaire qui vont nous intéresser, comment agréger le comportement des cellules pour rendre compte des observations que l'on souhaite faire dans un cadre radiobiologique, et comment propager les événements rapides sur de grandes durées de temps, ou étudier les phénomènes lents sur de brèves périodes.

Nous avons alors regardé ce qu'il existait comme modèles considérant ces caractéristiques. Pour les détailler nous allons commencer par décrire les règles structurelles des modèles existants, les types de modèles, puis nous regarderons les règles fonctionnelles, les différents processus biologiques qui ont été modélisés et enfin nous détaillerons quelques outils de modélisation appliqués aux systèmes qui nous intéressent.

III.2. MODELES DE POPULATION CELLULAIRE : DIFFERENTES APPROCHES

Dans son livre *Modeling and Simulation* de 1994, Bossel (Bossel, 1994 [9]) décrit qu'un modèle est élaboré pour répondre à une question précise, et par conséquent a pour but d'apporter une réponse à une question donnée. De fait, le modèle doit prendre en compte

les processus clés du système essentiels à l'étude de la problématique, en écartant un maximum de détails. La modélisation doit ainsi permettre d'obtenir une meilleure compréhension du système, en assurant le test d'hypothèses ou de scénarii, le réalisme ou la justesse d'idées ou de prédictions. Il convient de distinguer les modèles qualitatifs des modèles quantitatifs. Les premiers vont s'intéresser aux mécanismes du système, les seconds feront l'objet de l'étude mathématique du système. Avec le choix et l'élaboration du modèle se fera le choix de modèle continu ou discret pour les paramètres de temps, d'espace et d'état.

Il existe deux grandes catégories de modèles pour simuler les systèmes complexes, les modèles à population et les modèles individu-centrés, appelés également modèles à agents ou modèles individu-centrés. Les systèmes biologiques ne font pas exception, et peuvent être décrits par des modèles appartenant à ces deux catégories.

Les modèles à population décrivent un comportement global à l'échelle de la population ou du système dans son ensemble. Ce sont des modèles décrits par des équations mathématiques (équations différentielles, équations à dérivée partielle, équations non-linéaires, etc...) ou des systèmes d'équations.

De par leur nature, ces modèles décrivent un comportement global, à l'échelle de la population uniquement. Ils ne permettent pas d'observer le comportement des individus qui composent la population indépendamment les uns des autres. Ils ne sont pas adaptés à l'étude des comportements des éléments qui composent la population ou de sous-groupes, même si parfois le changement d'échelle des paramètres permet une observation à différents niveaux. Par exemple, le modèle MAED (Modèle pour l'analyse et la demande d'énergie) de l'Agence Internationale de l'Energie Atomique (IAEA, [10]) permet d'évaluer la demande finale en énergie au cours de l'année, et permet une observation au niveau d'un foyer, d'une agglomération, d'une région ou d'un pays selon l'échelle des paramètres utilisés, avec la possibilité d'étudier séparément la demande industrielle, la demande des ménages, etc...

Pour observer finement les comportements individuels, il faut se tourner vers les modèles individu-centrés.

Les modèles individu-centrés sont décrits par des règles qui s'appliquent à chaque individu et régissent leur comportement propre. Ces règles peuvent être simples ou complexes. L'observation de comportements mésoscopiques et macroscopiques reste possible car ils émergent de l'agrégation des comportements microscopiques.

Sous l'appellation comportement microscopique, c'est un comportement unitaire qui est désigné, c'est-à-dire le comportement d'un individu (une particule par exemple), d'une unité de surface (plus petite surface d'étude d'un modèle climatique par exemple) ou de volume (plus petit volume d'étude d'un modèle hydrodynamique par exemple). A l'inverse, le comportement macroscopique désigne un comportement global à l'échelle de la population dans son ensemble ou couvrant tout l'espace. Entre ces deux échelles, nous parlons de comportement mésoscopique, et nous nous intéressons alors à ce qu'il se passe pour un groupe d'individus ou dans une portion de l'espace. Par définition, les modèles à population permettent l'étude des comportements macroscopiques, voire mésoscopiques.

Dans la description de notre objectif en radiobiologie, nous avons identifié le besoin de pouvoir gérer les cellules individuellement, chacune agissant de façon autonome suivant ses propres règles, ayant chacune son propre comportement et possédant son propre patrimoine. C'est donc vers les modèles individu-centrés que nous nous sommes naturellement tournés.

Nous avons regardé quelques modèles individu-centrés couramment employés pour la modélisation de systèmes biologiques. Ces derniers sont des systèmes dynamiques complexes dont le comportement résulte d'interactions de nombreuses entités en interaction à échelles de temps et d'espace différentes, communiquant entre elles et avec leur environnement. Les systèmes biologiques apparaissent comme des sujets de choix pour l'étude à travers des modèles individu-centrés, ceux-ci présentant une grande analogie de fonctionnement avec le vivant, considérant des entités discrètes qui obéissent à des règles communes ou propres à chacune. Ces modèles individu-centrés permettent en plus de voir apparaître des comportements complexes, à des niveaux mésoscopique et macroscopique, sans qu'ils aient été décrits dans les règles de fonctionnement microscopiques qui régissent le fonctionnement des individus. Dès lors, ils sont apparus comme un choix évident pour l'étude de notre problématique.

Dans ces modèles individu-centrés, les entités se caractérisent par un état pouvant évoluer dans le temps en réponse à une stimulation extérieure, ou en réponse à une modification de l'environnement. L'état d'un individu appartient à l'espace d'états défini en amont de la modélisation. Les modifications de l'environnement ou les stimulations extérieures modélisées peuvent correspondre à des champs de force (gravité, champs électriques, gradients électrochimiques, pression...) ou des interactions cellulaires (contact et échange d'informations).

Selon les variables de temps, d'espace et d'état, ces modèles peuvent être discrets ou continus, avec n'importe quelles combinaisons.

Avant de détailler les modèles individu-centrés qui ont retenu notre attention, nous allons commencer par une courte présentation de quelques modèles à population utilisés pour la modélisation des systèmes biologiques.

III.2.1. MODELISATION A POPULATION POUR LES PROCESSUS BIOLOGIQUES

Nous avons décrit dans le chapitre précédent la croissance cellulaire en milieu de culture et avons vu qu'elle se déroulait en quatre phases : la latence (Figure III.-1), la phase exponentielle, ((Figure III.2-3), la phase stationnaire (Figure III.-5) et la phase de mortalité ((Figure III.2-6) pour chacune desquelles la durée dépend des conditions de culture et de la cellule cultivée. Entre ces grandes phases, on peut considérer des périodes de transition qui s'intercalent entre la phase de latence et la phase exponentielle, puis entre la phase exponentielle et la phase stationnaire qui sont respectivement la phase d'accélération (Figure III.2-2) et la phase de décélération (Figure III.2-4).

Whiting et Buchanan ont proposé en 1993 (Whiting et Buchanan, 1993 [11]) de classer les modèles à population destinés à la modélisation de cette cinétique de croissance bactérienne en trois catégories. Les modèles dits « primaires » ont pour objet de modéliser la cinétique de la croissance bactérienne, les modèles « secondaires » ont pour objet d'étude l'influence des caractéristiques de l'environnement sur les paramètres de la cinétique de croissance. Les modèles « tertiaires » réalisent des cinétiques de croissance microbiennes tout en prenant en compte l'effet des conditions de culture, c'est-à-dire qui appliquent les règles des modèles secondaires sur la simulation de cinétiques de croissance par des modèles primaires.

Une revue de ces derniers a été réalisée par Thomas Ross en 2003 (Ross *et al.*, 2003 [12]). Ils sont utilisés dans l'industrie pour qualifier les processus de fabrication, et évaluer les conditions de préparation, de conditionnement et de stockage des produits alimentaires, en réduisant les risques sanitaires liés à la multiplication des bactéries. Ils concernent essentiellement les microorganismes pathogènes pour l'homme.

Les modèles concernés par la classification proposée par Whiting et Buchanan ne concernent la cinétique que jusqu'à la phase stationnaire incluse. Pour la phase de déclin, il faut faire appel à d'autres modèles qui sont les modèles d'inactivation ou modèles de survie.

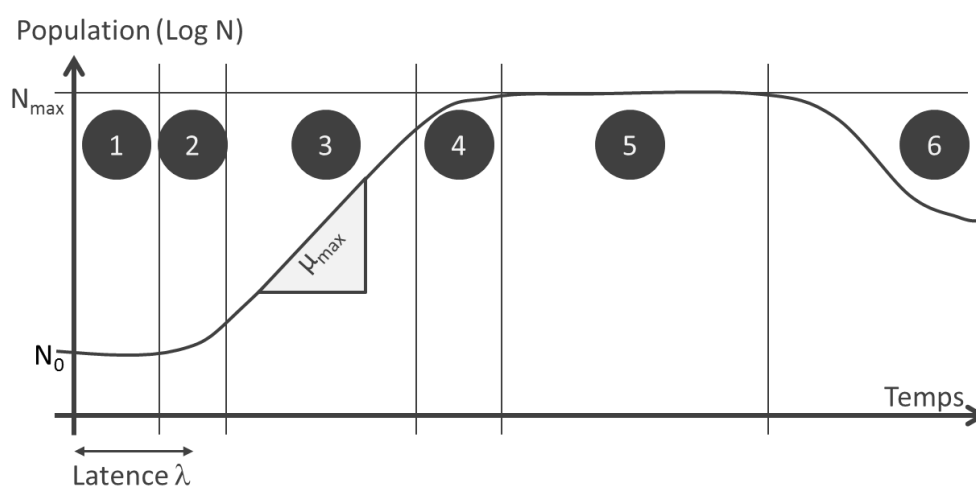


Figure III.2 : Les phases de la croissance lors d'une culture cellulaire en batch, inspiré de Buchanan (Buchanan, 1918 [4]).

La culture microbienne étant quelque chose de facile à réaliser en laboratoire et maîtrisée depuis la fin du 19^{ème} siècle, c'est sur les résultats expérimentaux que se sont développés les modèles de cinétique de croissance. En 1918, Buchanan s'est intéressé à sa description mathématique. Il a décrit chacune de ces phases indépendamment les unes des autres et a présenté un modèle exponentiel pour la phase exponentielle de croissance (Buchanan, 1918 [4]). Ce premier modèle est aussi connu sous le nom de modèle de Malthus du nom de Thomas Malthus qui l'a défini à la fin du 18^{ème} siècle pour décrire la croissance d'une population quelconque (Malthus, 1798 [13]; 1809 [14]). Il indique qu'une population croît à un taux fixe en échelle logarithmique, qui est proportionnel à sa taille. Dans le cas des bactéries, ce taux, appelé μ_{max} , caractérise le taux de croissance maximum obtenu dans les conditions expérimentales. Ce modèle s'écrit sous la forme

$$\begin{cases} \frac{dN(t)}{dt} = \mu_{max} \cdot N(t) \\ N(0) = N_0 \end{cases}$$

Où

- $N(t)$, le nombre de cellules à l'instant t ;
- N_0 , l'effectif initial.

Dont la solution s'écrit

$$N(t) = N_0 \cdot e^{\mu_{max} \cdot t}$$

Ce modèle qui ne peut rendre compte que de la phase exponentielle de croissance a été étendu en un modèle linéaire à trois phases pour permettre en plus, le rendu de la phase de latence et de la phase stationnaire (Buchanan *et al.*, 1997 [15]). À cette fin, les auteurs utilisent une durée fixe pour désigner la phase de latence λ durant laquelle la population reste constante à sa valeur initiale N_0 , et un effectif cellulaire maximal N_{max} , également fixe, qui permet de maintenir la population à une taille constante lors de la phase stationnaire. L'absence de transitions entre ces phases, et le fonctionnement avec des limites, reflètent mal la réalité.

Le modèle exponentiel produit une population dont la taille croît indéfiniment, et le modèle linéaire à trois phases stoppe la croissance à une taille de population précise. L'adjonction d'une fonction de freinage permet de reproduire la transition entre la phase exponentielle et la phase stationnaire en infléchissant la croissance. La fonction de freinage trouvée dans la littérature peut être de deux types : linéaire ou exponentiel.

Le premier cas, la croissance exponentielle avec freinage linéaire, est connu sous le nom de modèle logistique. Il a été défini par Pierre-François Verhulst en 1838 (Verhulst, 1838 [16]) et s'écrit

$$\begin{cases} \frac{dN(t)}{dt} = \mu_{max} \cdot N(t) \left(1 - \frac{N(t)}{N_{max}}\right) \\ N(0) = N_0 \end{cases}$$

Dont la solution s'écrit

$$N(t) = \frac{N_{max}}{1 + e^{-\mu_{max} t} \left(\frac{N_{max} - N_0}{N_0}\right)}$$

Ce modèle est présenté dans la littérature avec K en lieu et place de N_{max} , et K est défini comme la capacité de charge maximale d'un milieu, c'est-à-dire le nombre maximum d'êtres vivants que peut contenir un milieu. Le nombre de cellules maximal qui peut être maintenu dans le milieu reste ainsi une limite explicite dans l'énoncé de ce modèle.

Le second cas, connu sous le nom de modèle de Gompertz, a été défini par Benjamin Gompertz en 1825 (Gompertz, 1825 [17]). Il offre une plus grande finesse dans la gestion des

transitions entre les phases en considérant que le taux de croissance suit une exponentielle inverse au fur et à mesure de la croissance de la population. Ce modèle s'écrit

$$\begin{cases} \frac{dN(t)}{dt} = \mu_{max} N(t) e^{-\alpha t} \\ N(0) = N_0 \end{cases}$$

Et sa solution se note

$$N(t) = N_0 \frac{e^{-\frac{\mu_{max} e^{-\alpha t}}{\alpha}}}{e^{-\frac{\mu_{max}}{\alpha}}}$$

Si celui-ci permet de reproduire une cinétique de croissance correspondant aux données expérimentales pour les phases exponentielle et stationnaire, il utilise dans son expression un paramètre, dont l'interprétation en termes de données ayant une signification biologique peut s'avérer difficile, le paramètre α . La taille maximale de la population n'apparaît plus explicitement, mais à travers cette écriture, cette limite survient lorsque le temps qui correspond à l'apparition de l'effectif maximum approche.

Des réécritures de l'équation de ce modèle ont été proposées en reprenant les paramètres N_0 , μ_{max} , N_{max} et λ de la cinétique de croissance bactérienne (Zwietering *et al.*, 1990 [18]). La solution devient alors

$$N(t) = N_0 e^{A \cdot e^{-\left(\frac{\mu_{max} \cdot e}{A} \cdot (\lambda - t) + 1\right)}}$$

Où $A = \ln(N_{max}) - \ln(N_0)$

Ce modèle ne permet pas de satisfaire aux exigences de l'industrie agro-alimentaire qui s'intéresse au contrôle de la croissance avant que la phase exponentielle ne commence, pour assurer la qualité de ses produits, il faut en effet que les bactéries ne se multiplient pas et ne dégradent pas le produit ou ne présentent pas de risque sanitaire. Il était donc nécessaire d'apporter un modèle permettant de rendre compte de la phase de latence, sans la faire apparaître explicitement dans son écriture.

József Baranyi a proposé en 1993 un nouveau modèle en ce sens. Son approche est basée sur l'utilisation de paramètres ayant une signification biologique (Baranyi *et al.*, 1993 [19]) et il associe ainsi la phase de latence à un paramètre qu'il décrit comme l'état métabolique des cellules. Il considère que la phase stationnaire dépend également de ce paramètre. Cet état métabolique est lié, au cours de la phase de latence, à un temps d'adaptation des cellules à leur milieu, et au cours de la phase stationnaire, il est lié à un épuisement des ressources disponibles. Ce modèle s'écrit sous la forme

$$\begin{cases} \frac{dN(t)}{dt} = \frac{z_1(t)}{K_z + z_1(t)} \cdot \mu_{max} \left(1 - \left(\frac{N(t)}{N_{max}}\right)^m\right) \cdot N(t) \\ N(0) = N_0 \end{cases}$$

Où $\frac{dz_1(t)}{dt} = \nu z_1(t)$ décrit la temporisation de la croissance, c'est à dire la phase de latence et où la fonction de freinage, $1 - \left(\frac{N(t)}{N_{max}}\right)^m$, ressemble fortement à celle employée dans le modèle logistique.

La solution de ce modèle s'écrit

$$y(t) = y_0 + \mu_{max}t + \frac{1}{\mu_{max}} \ln(e^{-\nu \cdot t} + e^{h_0} - e^{-\nu \cdot t - h_0}) - \frac{1}{m} \ln \left(1 + \frac{e^{m\mu_{max}t + \frac{1}{\mu_{max}} \ln(e^{-\nu \cdot t} + e^{h_0} - e^{-\nu \cdot t - h_0})} - 1}{e^{m(y_{max} - y_0)}} \right)$$

Où :

- $y(t) = \ln(N(t))$
- $y_0 = \ln(N_0)$
- m , le paramètre de courbure qui caractérise la transition depuis la phase exponentielle vers la phase stationnaire
- ν , le paramètre de courbure qui caractérise la transition depuis la phase de latence vers la phase exponentielle
- h_0 , un paramètre sans dimension qui représente l'état physiologique initial des cellules de l'inoculum ; la durée de la phase de latence λ est calculée par la formule $\lambda = h_0 / \mu_{max}$.

Laurent Delhalle a réalisé en 2012 un panorama des modèles primaires, des modèles secondaires et des modèles tertiaires (Delhalle *et al.*, 2012 [20]). Une conclusion intéressante de ce document dit qu'il est actuellement difficile de définir si un modèle de croissance est plus approprié que les autres pour simuler la croissance microbienne, les modèles les plus simples pouvant suffire pour décrire les paramètres fondamentaux de la croissance.

Pour modéliser finement la cinétique de croissance microbienne, s'il apparaît assez peu discutable que la phase exponentielle est la principale composante, obtenir les autres phases nécessite l'élaboration d'un nouveau modèle qui présente la possibilité d'avoir une phase de croissance très lente d'abord, la phase exponentielle ensuite, puis la phase stationnaire pour terminer. Écrire un tel modèle avec les phases de transition impose de définir une fonction de temporisation qui permet de retarder la survenue de la phase exponentielle, et d'une fonction de freinage qui permet de réduire le rythme de la croissance, pour donner la phase stationnaire.

III.2.2. MODELISATION INDIVIDU-CENTREE POUR LES SYSTEMES BIOLOGIQUES

Nous avons vu que les modèles individu-centrés sont décrits par des règles qui s'appliquent à chaque individu, et régissent leur comportement propre. Ces règles peuvent être simples

ou complexes et permettent d'observer des comportements individuels, qui peuvent être simples ou complexes. Rajoutons la notion de « superindividu », qui permet de décrire par une seule entité plusieurs individus en présentant des caractéristiques moyennes (Scheffer, 1995 [21]). Le superindividu est parfois lié à l'approche mésoscopique, et constitue un compromis entre un modèle à population et un modèle individu-centré trop précis.

Parmi les modèles individu-centrés couramment employés pour la modélisation de systèmes biologiques, nous avons retenu les Automates Cellulaires, le Modèle de Potts, le Système Lattice-Gas, et les Systèmes Multi-Agents. Nous présentons pour chacun son historique avant de décrire ses caractéristiques et quelques exemples d'utilisation. Il existe plusieurs excellents documents qui référencent des utilisations de ces modèles pour la modélisation de systèmes biologiques, trois ont particulièrement retenu notre attention et ont servi à l'élaboration de la présente section (Ermentrout *et al.*, 1993 [22] ; Alber *et al.*, 2003 [23] ; Deutsch et Dormann 2004, [24]).

III.2.2.1. AUTOMATES CELLULAIRES

Le premier type de modèle individu-centré qui a retenu notre attention est le modèle des automates cellulaires (AC).

Un automate cellulaire est un système de cellules interagissant localement de manière simple et manifestant un comportement global complexe. Il se présente sous forme d'une grille à n dimensions divisée en « cellules ». Chaque cellule peut prendre à un instant t donné un état pris parmi un ensemble de valeurs, fini ou non. L'état d'une cellule dans le temps évolue selon les règles fonctionnelles. Dans les règles fonctionnelles on retrouve les règles de transition, regroupées dans la fonction de transition, et d'autres propriétés de la dynamique de l'automate cellulaire (déterminisme, probabilités, mise à jour, ...).

À chaque nouvelle unité de temps, l'application des règles structurelles et fonctionnelles à chaque cellule de la grille donne une nouvelle génération de cellules.

Le comportement global de l'automate s'observe en étudiant la totalité de la grille, alors que les règles fonctionnelles ne décrivent que les événements qui concernent uniquement chaque cellule séparément. L'observation se fait au niveau macroscopique, voire au niveau microscopique, mais les règles fonctionnelles ne sont appliquées qu'au niveau microscopique.

L'Encyclopédie Universalis présente les automates cellulaires comme le fruit de la recherche sur la vie artificielle, et l'exploration de règles simples permettant l'émergence de comportements complexes à plus haut niveau. Les exemples souvent cités sont d'une part les colonies d'insectes sociaux (fourmis, abeilles, termites) dans lesquels l'observation des individus isolés ne permet pas de deviner le comportement global et l'organisation générale de la colonie, et d'autre part le cerveau dans lequel l'observation d'un neurone isolé et de ses impulsions électriques ne donne pas un aperçu d'une forme de pensée. Dans ces cas, le comportement global complexe découle de l'ensemble des comportements propres à chaque individu, sans que quelque chose dans le comportement de chacun ne préside au comportement global.

Historiquement, on doit l'invention des Automates Cellulaires à John Von Neumann qui cherchait à savoir s'il était possible de concevoir une machine capable de s'autoreproduire. Les difficultés techniques pour la réalisation pratique d'un tel prototype le poussèrent à utiliser, sous la suggestion de Stanislaw Ulam, un système plus abstrait, les « objets géométriques définis de façon récursive ». Ces objets produisent, à partir de règles simples, des motifs dont le comportement semble évoluer dans un monde qui leur est propre, doté de ses propres lois, se propageant ou disparaissant. Ce monde est constitué d'une matrice infinie dont les cellules peuvent être actives ou inactives. Leur devenir à chaque instant dépend de la configuration de cellules avoisinantes.

Von Neumann exploita cette idée et la formalisa pour donner naissance au concept d'Automate Cellulaire. Il réussit, avec cet outil, à résoudre son problème d'autoreproduction en proposant en 1952 un automate utilisant 200.000 cellules, appartenant à 29 types différents, dont l'état dépend à tout moment de l'état des 4 cellules adjacentes disposées aux points cardinaux. Ce système de voisinage gardera le nom de voisinage de Von Neumann.

Le terme d'Automate Cellulaire ne sera adopté qu'en 1966 avec la parution de l'ouvrage *Theory of self reproducing automata* (Von Neumann, 1966 [25]).

Les Automates Cellulaires ont été popularisés par le Jeu de la Vie (Game of Life) de John Conway (voir pour le paragraphe IV.1.2.2, page 118 pour plus de détails) qui fut décrit dans un article de Martin Gardner (Gardner, 1970 [26]).

La définition originelle de l'automate cellulaire décrivait les interactions cellule-cellule par contact, dans un espace bidimensionnel régulier, avec un nombre fini d'états discrets. Les règles de transition définissent l'évolution de l'état d'une cellule en fonction de son état et des interactions dans lesquelles elle est impliquée. L'ensemble de ces règles forme la fonction de transition. Dans un automate cellulaire « classique », la fonction de transition est appliquée à l'identique, sur chaque cellule, à chaque pas de temps.

La définition des automates cellulaires a été enrichie et la description d'un automate cellulaire se fait maintenant par la caractérisation :

- Des règles structurelles
 - o La forme et la gestion de l'espace :
 - Nombre de dimensions (1 ou plus),
 - Espace continu ou discret,
 - Forme du maillage,
 - Comportement aux bords de la matrice (périodique, miroir, etc...);
 - o Le type et le nombre d'états que peut prendre chaque cellule :
 - États continus ou discrets,
 - Nombre d'états possibles;
 - o Le type de voisinage pris en compte à l'application de la fonction de transition sur chaque cellule, c'est-à-dire les autres mailles de l'espace dont l'état peut influencer le résultat du calcul de la fonction de transition (Figure III.3).

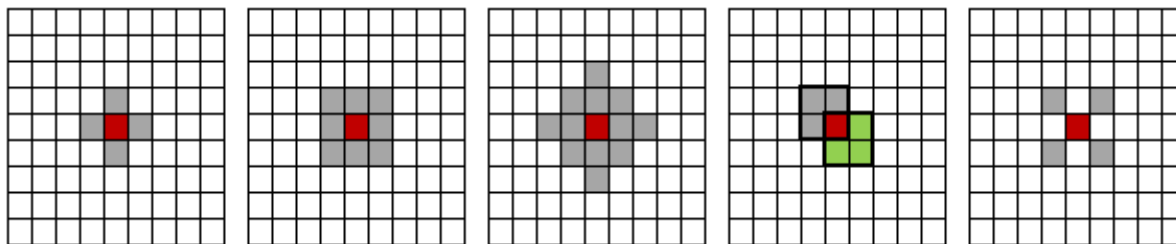


Figure III.3 : Quelques exemples de voisinages employés dans les automates cellulaires. La cellule active est en noir, les voisines prises en considération dans l'application de la fonction de transition apparaissent grisées. De gauche à droite, voisinage de Von Neumann, voisinage de Moore, voisinage 2D-radial, voisinage de Margolus (deux groupes de cellules sont pris alternativement comme voisinage pour le calcul de la fonction de transition) et voisinage diagonal.

- Des règles fonctionnelles
 - o La fonction de transition, c'est l'ensemble des règles de transition qui peuvent exister au sein de l'automate cellulaire. Elle peut être
 - Appliquée à toutes les cellules au même pas de temps ou non, on parle respectivement d'automate cellulaire synchrone ou d'automate cellulaire asynchrone,
 - Composée des mêmes règles pour toutes les cellules (automate cellulaire homogène) ou présenter des règles différentes selon les cellules (automate cellulaire hétérogène) ;
 - o Le type de règles de transition appliquées :
 - Règles déterministes ou règles probabilistes, on parle respectivement d'automate cellulaire déterministe ou stochastique.

Un automate cellulaire fonctionne avec un temps discret. La valeur d'état de chaque cellule à chaque pas de temps dépend de son état et de celui de ses voisines. La définition originelle d'un automate cellulaire décrit le cas de cellules synchrones. Leur état à $t+1$ est calculé en fonction de leur état et de celui de leur voisine au temps t . Lorsque l'état de toutes les cellules à $t+1$ a été calculé, on passe au pas de temps suivant et on calcule l'état de toutes les cellules à $t+2$. Et ainsi de suite. Lorsqu'on considère un automate cellulaire asynchrone, on peut le faire de deux façons. Dans la première, les cellules sont mises à jour une par une au fur et à mesure du calcul de leur état jusqu'à ce que toutes les cellules soient mise à jour, on passe alors au pas de temps suivant et les cellules sont resynchronisées, c'est le cas asynchrone-synchrone. La seconde façon, asynchrone, fait disparaître la notion de pas de temps en effectuant la mise à jour et l'application du nouvel état à une cellule prise à chaque fois au hasard jusqu'à arrêt de la simulation.

On peut par exemple décrire un modèle simplifié de diffusion sur un automate cellulaire synchrone à deux dimensions dans un espace discret avec un maillage régulier et des conditions absorbantes aux bords. Les états sont discrets et au nombre de deux : soluté et solvant. Le voisinage employé est le voisinage de Von Neumann. La fonction de transition contient une règle qui peut s'écrire « si la cellule active a dans son voisinage une cellule qui contient du soluté, son état devient soluté, sinon son état est solvant » (Figure III.4). L'automate cellulaire est synchrone et homogène.

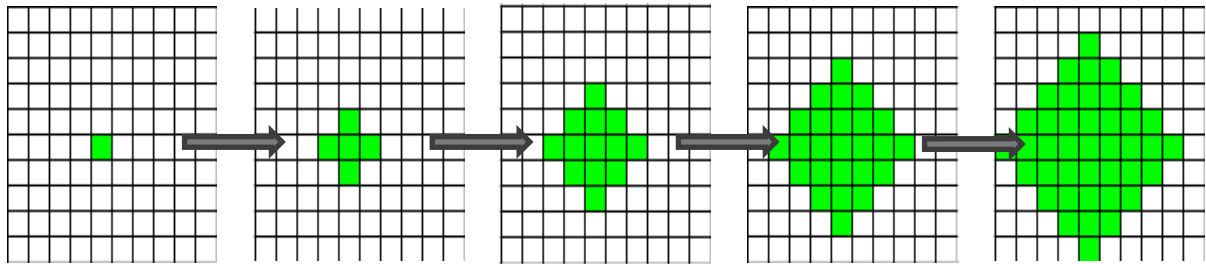


Figure III.4 : Un exemple d'automate cellulaire : un modèle simplifié de diffusion.

Les automates cellulaires ont été utilisés pour modéliser des systèmes variés : machine de Turing (Albert et Culik, 1987 [27] ; Cook, 2004 [28]), des systèmes biologiques tels que par exemple la croissance tumorale (Dormann et Deutsch, 2002 [29]), la stratification d'un épithélium (Maheswaran *et al.*, 2007 [30]), l'embryogenèse (Laforge, 2005 [31]), la morphogénèse (Markus *et al.*, 1998 [32]), la formation des motifs caractéristiques d'animaux (motifs sur la peau des mammifères, des reptiles ou des poissons (Cocho *et al.*, 1986 [33] ; motifs sur la coquille des mollusques (Kush et Markus, 1995 [34]), Figure III.5), la propagation de feux de forêts (Karafyllidis et Thanailakis, 1997 [35] ; Alexandridis *et al.*, 2008 [36]), la propagation d'épidémies (White *et al.*, 2007 [37] ; Ching Fu, 2002 [38])... On trouvera plusieurs exemples dans le livre *Cellular Automaton Modeling of Biological Pattern Formation* de Deutsch et Dormann (Deutsch et Dormann 2004, [24]).

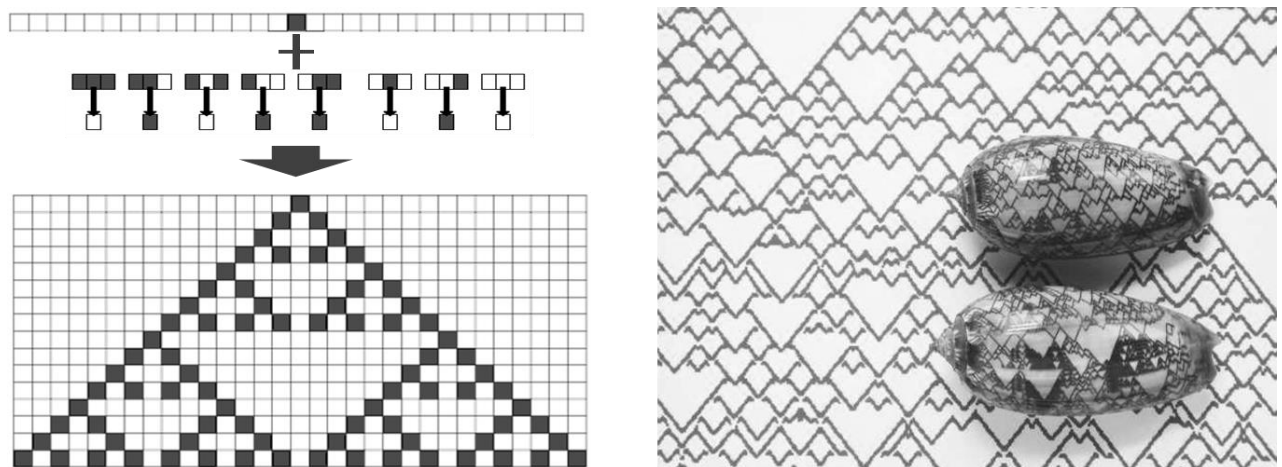


Figure III.5 : Exemple de mise en œuvre d'un automate cellulaire pour reproduire un motif rencontré dans la nature, le motif de la coquille d'*Oliva porphyra* [39].

III.2.2.2. MODELE DE POTTS

Le second type de modèle que nous présentons est le modèle de Potts.

Le modèle de Potts est parfois décrit comme un automate cellulaire modifié. Il consiste en effet en une grille régulière de cellules présentant une valeur décrivant un état. La différence majeure par rapport à l'automate cellulaire est que l'individu peut couvrir plusieurs mailles, là où un individu de l'automate cellulaire ne peut s'inscrire que dans une seule maille. Dans un modèle de Potts, un individu est représenté par une région de la matrice, les mailles

disposent d'une « étiquette » qui leur permet de savoir à quel individu elles appartiennent (Figure III.6).

1	1	1	1	2	2	4	4	4	4
1	1	1	2	2	2	2	4	4	4
1	1	2	2	2	2	2	4	4	4
1	1	2	2	2	2	4	4	4	4
3	3	3	2	5	5	5	4	4	4
3	3	3	3	5	5	5	5	5	7
3	3	3	5	5	5	5	5	7	7
3	3	3	5	5	5	5	7	7	7
3	3	6	6	6	6	6	7	7	7
3	3	6	6	6	6	6	7	7	7

Figure III.6 : Exemple d'un modèle de Potts, 7 cellules sont représentées sur une grille de 10 cellules par 10. Chaque cellule est identifiée par un numéro qui permet d'identifier les cases du maillage qui lui sont associées.

Le modèle de Potts a été développé pour étudier les interactions de spins dans les réseaux cristallins (Potts, 1952 [40]). Le modèle de Potts consistait alors en une généralisation avec plus de deux valeurs de spin du modèle proposé en 1920 par Enst Ising dans sa thèse pour l'étude des transitions de phase des moments magnétiques dans les matériaux magnétiques. Dans ce modèle, l'ordre de mise à jour des sites suit une distribution de Monte-Carlo. Le temps est mesuré en Monte-Carlo Steps ou MCS.

Le principe de ce modèle est de permettre de diminuer l'énergie du système en minimisant l'énergie qui existe entre les « grains ». Optimiser les limites de ces grains diminue les contraintes qui existent à leur surface et donc l'énergie globale du système. Ce modèle a trouvé son utilisation, dans l'étude des systèmes biologiques, pour tous les processus d'organisation spontanée. Les cellules ont en effet un volume fini, leur surface est mobile selon les forces extérieures, et l'auto-organisation permet d'équilibrer les tensions subies par chacune (Figure III.7).

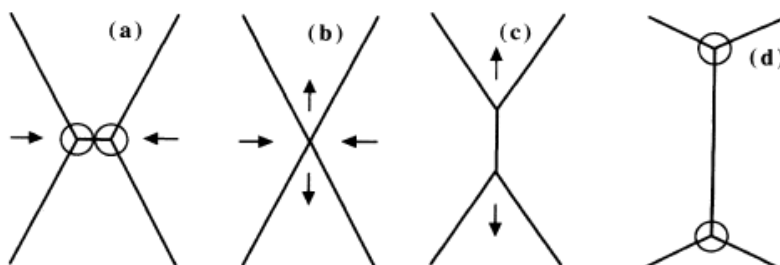


Figure III.7 : Exemple de réarrangement observé qui permet de diminuer la tension exercée à la surface de chaque cellule. Optimiser les contraintes exercées sur chaque cellule permet ainsi de diminuer l'énergie globale du système [41].

Il a ainsi été utilisé pour étudier la réorganisation des tissus par auto-organisation des cellules selon les affinités entre elles (Graner et Glazier, 1992 [42] ; Glazier et Graner 1993

[41] ; Figure III.8) ou selon la contrainte d'un champ externe (Jiang *et al.*, [43]), et la morphogénèse (Savil et Hogeweg, 1997 [44]).

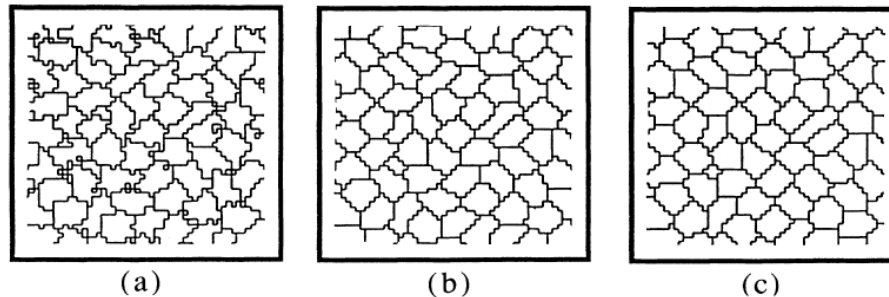


Figure III.8 : Exemple d'un réarrangement cellulaire réalisé sur un modèle de Potts, les cellules disposées aléatoirement au départ réorganisent leur membrane pour diminuer chacune la tension qui s'exerce dessus. Cette réorganisation à grande échelle permet d'abaisser l'énergie du système [41].

Ce modèle est parfois décrit comme Cellular Potts Model ou CPM (Alber *et al.*, 2003 [23]).

III.2.2.3. SYSTEME LATTICE-GAS

Le système Lattice-Gas est parfois traduit par Système de Particules en interaction.

Les systèmes Lattice-Gas sont constitués d'une matrice régulière dans laquelle chaque maille présente plusieurs sites, chaque site correspond à une direction, on parle de canal. Dans la description de ce modèle, le terme couramment employé n'est pas maille, mais node (Figure III.9). Une particule placée dans ce réseau se déplace de node en node, en passant dans les canaux correspondants à sa direction. Chaque particule embarque l'information de sa vitesse pour savoir de combien de nodes elle se déplace par pas de temps.

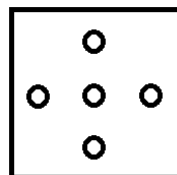


Figure III.9 : Exemple d'un node à cinq canaux, les canaux en haut, à droite, en bas et à gauche correspondant respectivement aux directions "vers le haut", "vers la droite", "vers le bas" et "vers la gauche". Au centre on trouve un canal neutre pour une particule dont la vitesse est nulle.

Ceci permet de gérer les collisions entre les particules, plusieurs particules pouvant occuper le même node si elles ont des directions différentes, et deux particules se trouvant face à face, sur deux nodes distincts, peuvent se confronter pour déterminer une nouvelle trajectoire (Figure III.10).

Ce modèle dérive du modèle dit HPP, de ses auteurs Hardy, de Pazzis et Pomeau qui s'intéressaient à la mécanique des fluides (Hardy *et al.*, 1976 [45]). Le modèle Lattice-Gas a vraiment commencé à se faire connaître sous cette forme en 1986 avec les travaux de Frish, Hasslacher et Pomeau sur l'équation de Navier-Stokes (Frisch *et al.*, 1986 [46]).

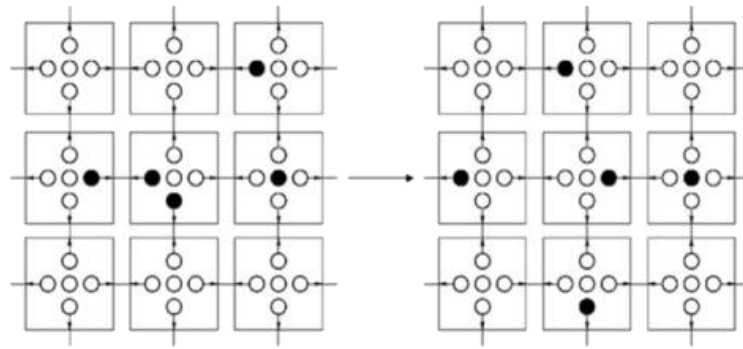


Figure III.10 : Exemple de l'évolution de 5 particules disposées dans un réseau Lattice-Gas. Deux pas de temps sont représentés.

La conception d'un modèle Lattice-Gas passe par la définition des propriétés des particules qui vont interagir. Parmi ces propriétés il va essentiellement s'agir de la masse, mais on peut aussi parler de la charge par exemple. En plus de ces caractéristiques intrinsèques, liées au type des particules, chacune se voit dotée d'une célérité. La masse, la charge et les autres propriétés que l'on peut envisager pour les particules ainsi représentées ont une influence lors la gestion des collisions entre particules. Ces collisions ont lieu lorsque les trajectoires des particules les amènent au contact l'une de l'autre, cela peut se produire par exemple lorsque deux particules sont dans les canaux en vis-à-vis de deux nodes adjacents. La vitesse de chacune et leurs propriétés vont permettre de calculer leur nouvelle trajectoire, par exemple, les deux particules vont rebondir l'une sur l'autre et repartir dans la direction opposée, ou elles vont ricocher et partir chacune dans une direction différente de leur direction incidente. Il faut donc concevoir les règles qui seront appliquées lors de ces collisions. Sans collision une particule se déplace selon sa direction, de node en node, d'un nombre de nodes dépendant de sa célérité à chaque pas de temps. Il est également important de définir le nombre de canaux présents par node, pour les trajectoires des particules, plus un node présentera de canaux, plus la gestion des directions pourra être riche, la version classique en comporte 5 (Figure III.).

La mise à jour d'un tel système se déroule en deux étapes, l'étape de propagation et l'étape de mouvement, ce qui permet d'éviter les situations de conflit. Le mouvement se déroule si nécessaire en plusieurs temps, il faut autant de temps que nécessaire pour satisfaire à la célérité d'une particule, à chaque temps, on vérifie si il y a collision ou non. L'ordre de mise à jour des nodes est aléatoire (Wolf-Gladrow, 2005 [47]).

Ce fonctionnement node à node permet de rapprocher ce modèle des automates cellulaires, et souvent le nom LGCA est employé pour Lattice-Gas Cellular Automata. La gestion des collisions est alors à rapprocher des règles de transition d'un automate cellulaire (Wolf-Gladrow, 2005 [47] ; Deutsch et Dormann, 2004, [24]). Une évolution récente du modèle LGCA consiste à considérer des modèles continus, en lieu et place des entités discrètes, on parle alors de LBCA pour Lattice Boltzmann Cellular Automata.

Ce modèle a essentiellement été utilisé pour la modélisation de systèmes physiques, en particulier la mécanique des fluides. Si cette dernière impose des matrices à plusieurs dimension pour donner des résultats concluants (Wolfram, 1986 [48] [49]), il existe quelques

implémentations intéressantes avec les systèmes biologiques dans l'hémodynamique (Ouared *et al.*, 2005 [50] ; Chopard *et al.*, 2007 [51]) ou la formation de métastases dans le développement cancéreux (Hatzikirou *et al.*, 2008 [52]).

III.2.2.4. SYSTEMES MULTI-AGENTS

Les systèmes multi-agents impliquent la modélisation d'agents en interaction. Il n'est pas nécessaire de modéliser un environnement, mais celui-ci peut être présent, même éventuellement sous la forme d'un agent à part entière. Les agents modélisés obéissent chacun à leurs règles propres et il peut exister plusieurs types d'agents.

Leur naissance se déroule dans les années 80 avec la convergence de l'algorithmique parallélisée, et les avancées en termes d'intelligence artificielle. Ils se destinent alors à l'étude des colonies d'insectes sociaux dont ils vont mimer le fonctionnement.

Un agent représente une entité quelconque qui se caractérise au moins par une autonomie partielle, c'est l'intelligence artificielle distribuée. Pour cela, un agent doit avoir deux éléments définis, d'une part un modèle cognitif qui lui permet la perception et l'évaluation de son environnement, et qui décrit également ses objectifs, et d'autre part un modèle de communication qui permet à l'agent d'avoir des échanges avec son environnement ou les autres agents.

Ce modèle suit un processus intégratif ce qui s'accompagne d'une hiérarchisation des individus. Les agents peuvent s'organiser de différentes façons : hiérarchies simples, confédérations, équipes, congrégations, holarchies, etc... Par exemple, dans les confédérations, un agent joue le rôle de délégué, il est l'interlocuteur des membres de la confédération et des délégués des autres confédérations.

Tous les agents d'un système peuvent être différents, et posséder chacun un modèle cognitif et un modèle de communication différent. On distingue les agents dits « réactifs » qui seront plus acteurs, et les agents dits « cognitifs » qui seront plus spectateurs ou décideurs.

Selon Grimm, les systèmes multi-agents répondent à deux raisons. La première qualifiée de « pragmatique » naît de la difficulté à étudier des phénomènes complexes à l'aide de modèles mathématiques globaux, citons par exemple le comportement d'un banc de poissons. La seconde raison est qualifiée de « paradigmatique », elle intervient lorsque les modèles mathématiques montrent leurs limites pour expliquer les phénomènes de régulation, d'émergence, de résilience, de persistance, etc... dans les écosystèmes (Grimm, 1999 [53]). Pour autant, il semble ne pas y avoir de définition consensuelle et précise d'un Système Multi-Agent et la validation d'un tel modèle est délicate, faute d'outils et de méthodologie (Amblard *et al.*, 2006 [54]; Amigoni et Schiaffonati, 2007 [55]), même si des travaux vont dans ce sens (Grimm *et al.*, 2006 [56]).

On trouve toutefois de nombreuses utilisations dans des domaines tels que l'écologie (Ginot et Le Page, 1998 [57] ; Brochier *et al.*, 2012 [58]), la finance (Lin et Chang, 2004 [59]) ou la résolution de problèmes complexes tels que le Voyageur de Commerce (Ilie et Badica, 2013 [60]). Dans les sciences du vivant, pour les échelles qui nous intéressent, nous citerons en exemples le repliement des protéines (Bortolussi *et al.*, 2005 [61] ; Orro *et al.*, 2005 [62]), le

système immunitaire (Bandini *et al.*, 2005 [63]) ou les fonctions cellulaires (Bosse *et al.*, 2005 [64]).

III.2.3. BILAN SUR LES TYPES DE MODELES

Ermentrout en 1993 indique que les modèles réalistes sont basés sur les lois physiques, qui sont issues de systèmes à grande échelle (Ermentrout, 1993 [22]). La modélisation fine a donc un avantage à être réalisée par une approche bottom-up. Néanmoins, si cette dernière apporte un niveau élevé de détails, cela se traduit par un coût important. Il faut de plus faire attention lors du développement du modèle, les règles décrites étant souvent liées au cadre expérimental, cela peut amener à produire des modèles lourds et obscurs. La simplification souvent employée consiste alors à remplacer les lois physiques par des séries de règles simples, faciles à implémenter et rapides à calculer. Ceci conduit à un intérêt croissant pour les systèmes à agents en sciences physiques et en sciences naturelles.

Des agents discrets dans un état donné évoluent dans le temps en fonction de l'état dans lequel ils sont et de ce qui les entoure, les autres agents et leur état ou le milieu dans lequel ils baignent. Il est nécessaire que chaque agent dispose de capacités d'interaction avec ses voisins ou son milieu, et qu'il dispose de capacités de réaction. Ces principes s'appliquent à tous les modèles que nous venons de décrire, il semble d'ailleurs exister une confusion sous le terme SMA qui ne désigne pas seulement le modèle des Systèmes Multi-Agents mais l'ensemble des modèles individu-centrés (Bousquet *et al.*, 2002 [65]). Hwang parle ainsi de modélisation basée sur les agents (ABM) pour généraliser la description des modèles faisant intervenir des agents discrets en interaction dans un environnement donné (Hwang *et al.*, 2009 [66]). De plus, les Systèmes Multi-Agents présentent souvent une hiérarchie entre les agents, ce qui ne nous apparaît pas cohérent avec les systèmes biologiques, et nous préférons les modèles dans lesquels les individus sont tous sur un pied d'égalité.

Les modèles de Potts et Lattice-Gas peuvent être considérés comme des automates cellulaires (Hwang *et al.*, 2009 [66]), aussi notre réflexion s'est-elle axée sur les automates cellulaires, mais en conservant en perspective leurs nombreuses possibilités de développement, nous ne nous limitons pas à leur définition initiale. Par la suite nous conserverons le terme d'Automate Cellulaire, mais nous le considérerons dans son sens large.

Ainsi, les automates cellulaires répondent au paradigme des systèmes complexes et constituent un outil bien adapté à l'étude de tels systèmes. En effet, au sein d'un automate cellulaire, la description de règles locales simples qui vont diriger le devenir de chaque cellule au fil du temps en fonction de ses voisines, va permettre, à travers les interactions entre les parties, de faire émerger des propriétés ou des caractéristiques qui ne pourraient être devinées *a priori* en considérant la description des cellules.

C'est un type de modèle facile à implémenter et de conception simple, mais les interactions et relations entre les individus peuvent être complexes. Dans un automate cellulaire il n'y a cependant pas de notion de hiérarchie et aucun individu n'a de privilèges par rapport aux autres, les individus sont tous équivalents et n'obéissent qu'à leurs règles propres. Tout comme dans le système vivant dont il s'inspire et dont il tire son nom : la cellule est l'entité

fonctionnelle, autonome, elle est isolée de ce qui l'entoure mais elle est capable d'interagir avec les cellules adjacentes ou le milieu dans lequel elle baigne (Figure III.11), par contact, ou avec des cellules distantes par la transmission d'informations via l'environnement (à travers le milieu, ou à travers d'autres cellules, Deutsch et Dormann, 2004 [24]).

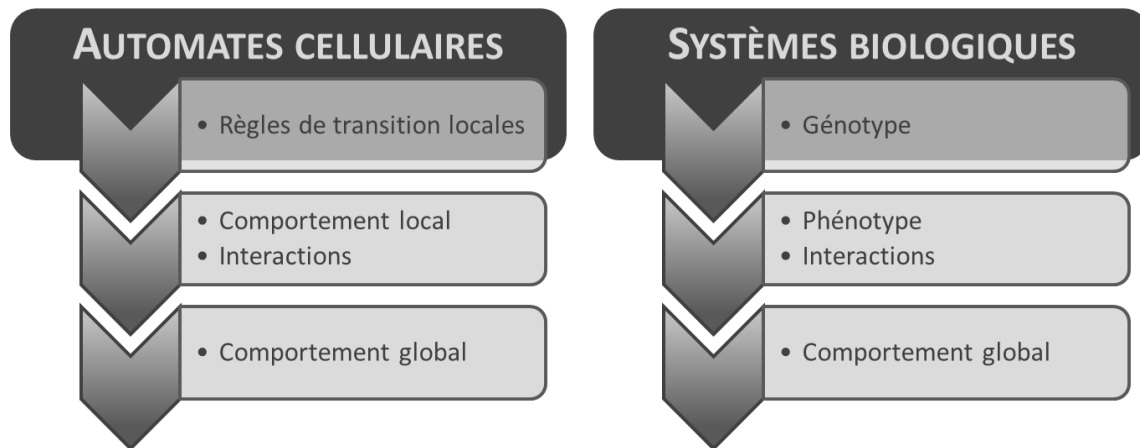


Figure III.11 : Analogie de fonctionnement entre les automates cellulaires et le système vivant duquel il s'inspire.

III.3. MODELES POUR LES PROCESSUS BIOLOGIQUES

Après avoir vu les types de modèles, les règles structurelles, nous explorons l'existant concernant la modélisation de quelques processus biologiques sans nous limiter à ceux implémentés sur des automates cellulaires (au sens large).

Si les types de modèles constituent les règles structurelles des modèles, les processus implémentés sont les règles fonctionnelles des modèles.

Plusieurs des modèles et outils trouvés sont élaborés dans un but pédagogique, pour permettre une meilleure compréhension de processus abstraits et permettre l'expérimentation de phénomènes inaccessibles en laboratoire (ou salle de Travaux Pratiques) car trop rapides, ou trop lents.

De nombreux processus biologiques ont été modélisés : embryogenèse, carcinogenèse, morphogenèse, etc... plusieurs de ces phénomènes ont déjà été abordés lorsque nous avons décrit pour chaque type de modèles quelques exemples d'application en biologie. Ermentrout (1993, [22]), Alber *et al.* (2003, [23]), Deutsch et Dormann (2004, [24]) par exemple ont chacun décrit plusieurs exemples de processus qui se déroulent à échelle d'une population de cellules, essentiellement chez des organismes complexes.

Il existe dans la littérature une classification de ces modèles selon qu'ils exploitent ou non trois processus biologiques élémentaires : la migration, la différenciation et la mort (Deutsch et Dormann, 2004 [24]; Hwang *et al.*, 2009 [66]). La migration consiste en le déplacement des cellules dans leur environnement, la différenciation décrit la transformation d'une cellule d'un type en un autre.

Ces trois processus sont irréductibles. Les cellules passent d'un état à l'autre. Une cellule ne peut pas être « à moitié morte » au cours d'un pas de temps. Des contraintes peuvent s'appliquer sur ces processus. Par exemple, la migration peut se faire, soit par déplacement aléatoire de la cellule dans la matrice, soit selon une direction indiquée par un champ de force, physique, chimique, ou physico-chimique. De même, la différenciation peut être réversible ou non, être liée à un facteur endogène à la cellule, ou lié à un facteur environnemental. La mort peut également être liée à des facteurs endogènes ou exogènes.

Notons que lorsque la mort est décrite, le devenir des cellules mortes n'est pas toujours un élément considéré. Soit la cellule qui meurt reste en place mais sans plus interagir avec ce qui l'entoure (environnement ou cellules voisines), ces modèles parlent de nécrose, soit la cellule disparaît purement et simplement sans autres traces, soit elle disparaît mais son contenu est recyclé dans l'environnement.

Ces trois processus permettent de décrire le fonctionnement des cellules à l'échelle d'un tissu ou d'une colonie de microorganismes, mais cela nous apparaît limitant pour décrire les échelles inférieures pour les processus subcellulaires, les processus biochimiques, mais également pour décrire les processus à échelles supérieures. Ceci sera particulièrement vrai si on parle des échelles temporelles, les processus biochimiques sont trop rapides pour apparaître dans les phénomènes de mort, de différenciation ou de migration, et l'évolution se déroule sur des temps infiniment longs par rapport à ces mêmes processus (Figure III.12).

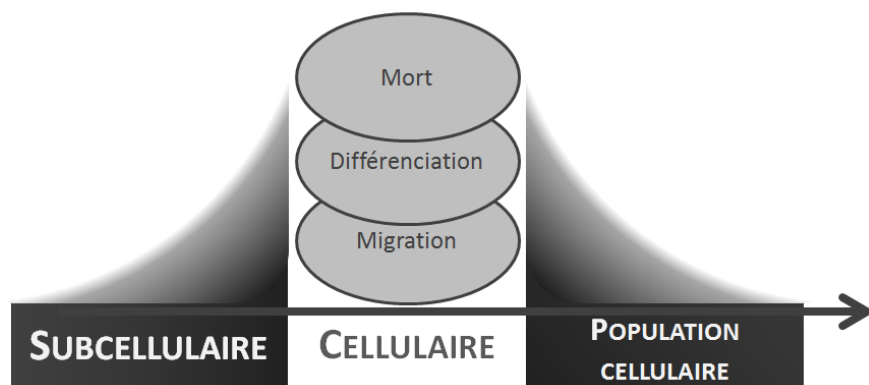


Figure III.12 : Les processus de migration, différenciation et mort ne concernent que l'échelle cellulaire et au phénomène d'adaptation.

Par contre, ces trois processus apparaissent comme les moyens qui vont permettre le développement d'une population donnée dans un environnement donné, et ainsi ils répondent bien à l'adaptation d'une population aux paramètres d'un environnement.

Nous proposons donc de classifier les modèles selon l'échelle temporelle adressée. Ce changement de perspective nous permet de passer outre les limitations de la classification basée sur la combinaison des phénomènes de mort, migration et différenciation pour placer des modèles qui traitent des échelles subcellulaires ou des phénomènes évolutifs.

Pour pouvoir nous intéresser au système vivant, quel qu'il soit, de la bactérie isolée à la colonie, du massif de cellules embryonnaires à l'individu, en passant par toutes les structures intermédiaires, et en considérant le devenir d'un individu au sein d'une population ainsi que

le développement d'une population, nous ne pouvons plus nous satisfaire de classer les modèles selon qu'ils considèrent une combinaison de trois processus élémentaires, mort, migration et différenciation.

En classifiant les modèles selon les échelles observées, en particulier en termes de temps nous voyons se dégager trois niveaux de détails, le premier niveau décrira les phénomènes très rapides au vu de la vie de l'organisme, et souvent à des échelles spatiales infiniment plus petites que celles de l'individu. Nous qualifierons ce premier niveau de « processus biochimiques ».

Le second niveau de cette classification adresse les phénomènes qui ont trait à l'individu, qui décrivent comment un individu réagit à un changement dans son environnement et à l'interaction avec les individus voisins. En termes de temps et d'espace, la vie de l'individu est au centre des considérations. Cette échelle correspond à la plupart des modèles biologiques abordés précédemment. Les réponses de l'individu doivent être adaptées aux simulations perçues sous peine de mener l'individu dans une impasse, aussi nous parlerons pour ce niveau d' « adaptation ».

Le troisième niveau de cette classification s'intéressera à tous les modèles dont l'objet de l'étude est infiniment grand par rapport à l'individu, en termes d'espace et de temps. Dans ces modèles, la vie d'un individu apparaît pratiquement comme un processus irréductible, ce qui importe est plus la trace que laisse l'individu que l'individu lui-même. On va observer le devenir d'une population, les changements qui s'opèrent en son sein, son développement ou sa régression, selon les conditions auxquelles elle est soumise. Ainsi, pour ce dernier niveau, nous choisirons le terme d' « évolution ».

Si l'on regarde le détail, un organisme s'adapte parce qu'il change sa façon d'interagir avec son environnement pour apporter une réponse adéquate. Pour un organisme de grande taille, on peut se limiter à considérer des membres ou des appendices comme « effecteurs », mais si on focalise notre attention sur l'échelle cellulaire, les effecteurs seront à l'échelle subcellulaire ou à l'échelle moléculaire. Il en est de même lors de l'évolution qui va se jouer en sélectionnant les individus les plus adaptés, si à l'échelle de l'organisme il peut s'agir de la modification ou de l'apparition d'un organe ou appendice, à l'échelle cellulaire, les changements que l'on observera se joueront à l'échelle moléculaire. Ainsi, l'échelle des processus biochimiques se révèle un élément transversal à l'adaptation et à l'évolution, et l'évolution elle-même dépend de processus d'adaptation.

Classer les modèles selon ces trois niveaux fait apparaître une hiérarchie qui peut être abordée par une approche intégrative ou réductionniste. Notons que les modèles dont l'objet est la modélisation de l'adaptation ou de l'évolution font appel aux mêmes entités. Ce sont les échelles de temps et les interactions modélisées qui font que le comportement est évolutif ou adaptatif.

III.3.1. PROCESSUS BIOCHIMIQUES

À travers ces modèles, ce sont des processus très rapides qui sont étudiés, de l'ordre de la microseconde à la milliseconde. En termes d'espace, les dimensions sont de l'ordre du nanomètre, voire moins.

Ces modèles sont utilisés pour modéliser les réactions chimiques à l'œuvre dans un système. Ce système peut être un tube à essai tout aussi bien qu'une cellule ou même qu'un groupe de cellules. Le type d'environnement n'a souvent pas d'importance, les dimensions spatiales du système étant largement plus importantes que les dimensions des entités modélisées.

Toutes ces réactions chimiques peuvent être structurées selon des chaînes de réaction et donner des voies métaboliques à part entière. Pour rappel, le métabolisme correspond à la somme de l'anabolisme et du catabolisme qui sont respectivement l'ensemble des réactions biochimiques mises en œuvre dans les processus de dégradation (consommation, absorption) et de construction (synthèse) nécessaires au bon fonctionnement d'une cellule, ainsi que les mécanismes de régulation qui contrôlent ces réactions.

Il intéresse les modèles dont l'objet est une séquence de réaction chimiques/biochimiques. Selon le nombre d'entités et de réactions en jeu, ils permettent de s'intéresser à un système simplifié ne présentant qu'une simple réaction ou système de transport par exemple, ou à un système plus élaboré tel qu'une voie métabolique complète avec des dizaines de réactions et d'entités en jeu.

Les entités modélisées vont représenter des espèces chimiques ou biochimiques capables de réagir entre elles en se conformant aux capacités des molécules biologiques représentées ou des réactions chimiques qui vont nécessiter la présence d'une quantité de réactifs pour pouvoir avoir lieu.

Les interactions peuvent être soumises à une probabilité et nécessiter une orientation de la rencontre, et la réaction peut nécessiter des conditions particulières (énergie, présence de cofacteurs, etc...) pour avoir lieu. Les entités modélisées sont placées dans un espace compartimenté ou non. Si l'espace est compartimenté, il faut définir pour chaque classe d'entité si elle est séquestrée ou non dans un compartiment particulier. Un modèle de diffusion est requis pour gérer le déplacement aléatoire des entités au sein de l'espace, ou à tout le moins un modèle permettant de gérer les déplacements des espèces chimiques vers leur lieu de réaction.

Nous en avons retenu trois exemples que nous présentons succinctement : SimCell, V-Cell et Snoopy.

A. SIMCELL

Ce modèle qui se veut une structure générale de simulation de la plupart des processus cellulaires est disponible en ligne ¹ (SimCell, [67]).

¹ <http://wishart.biology.ualberta.ca/SimCell/>

SimCell s'appuie sur les travaux de Wishart et al. (Wishart *et al.*, 2004 [68]) sur les Automates Cellulaires dans une variante que les auteurs appellent Automates Cellulaires Dynamiques. Ce type d'automate cellulaire consiste en une approche qualifiée de mésoscopique dans la mesure où une cellule de l'automate cellulaire présente au plus une seule macromolécule ou une multitude de petites molécules. Le reste des caractéristiques de l'automate cellulaire décrivent un automate cellulaire régulier à deux dimensions, avec un fonctionnement hétérogène, probabiliste et synchrone.

L'utilisation de SimCell se passe en trois temps (Figure III.13), d'abord la création du réseau à l'aide de l'interface graphique. Ensuite, la simulation du comportement de ce réseau à l'aide de l'automate cellulaire. Enfin, à tout moment de la simulation, des graphiques et des tableaux peuvent être générés pour présenter l'évolution des différentes entités impliquées dans le réseau modélisé.

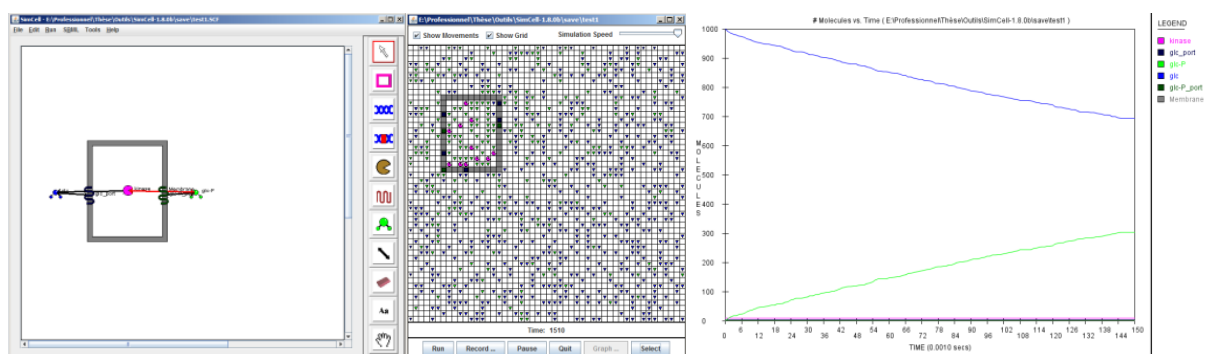


Figure III.13 : L'utilisation de SimCell se déroule en 3 temps, d'abord la création du réseau biochimique, la simulation du réseau par l'automate cellulaire, et le suivi du comportement des différentes espèces à l'aide de graphiques.

Pour construire le réseau, l'utilisateur dispose de six catégories différentes d'entités, chacune avec des propriétés et des possibilités de paramétrage différentes. Ces six catégories d'espèces chimiques se répartissent en deux groupes, les entités immobiles d'une part et les entités mobiles d'autre part.

Dans les entités immobiles on trouve :

- Membrane : Ce type d'objet servira à créer une membrane fermée, et donc, à créer un compartiment isolé dans l'espace. Les propriétés de cet objet permettent de définir sa taille ;
- ADN : Cet objet immobile figure une structure d'ADN double-brin sur laquelle des gènes ou des sites de régulation pourront être positionnés ;
- Gène : Ce type d'objet ne peut être positionné que sur une séquence d'ADN sur laquelle il va représenter un site d'interaction avec une macromolécule (Protéine ou ARN, avec un but de régulation positive ou négative) ou un gène.

Les entités mobiles sont :

- Petite molécule : Ce type d'objet servira à modéliser les petites molécules inorganiques (H_2O , CO_2 ...), les ions simples (Na^+ , Cl^- ...) ou composés (PO_4^{3-} , NH_4^+ ...), des petites molécules biochimiques simples. Chaque maille de l'automate cellulaire peut accueillir cent objets de ce type.

- Protéine ou ARN : Cet objet sert à créer des macromolécules qui ont une activité catalytique. La taille de cet objet correspond aux dimensions des cellules de l'automate, si bien qu'un seul objet ne peut être contenu par maille.
- Protéine membranaire : Cet objet sert à créer une protéine enchâssée dans une membrane. Ce type de protéine ne pourra donc être créé que si une membrane a préalablement été créée. Elle ne pourra pas se déplacer librement dans l'espace et restera confinée dans la membrane dans laquelle elle est placée.

Pour toutes ces entités on peut définir un certain nombre de propriétés :

- Nom ;
- Nombre ;
- perméabilité membranaire (sauf pour les protéines membranaires) ;
- répartition initiale (uniforme,...) ;
- taux de création/disparition

Les auteurs considèrent qu'une cellule de l'automate représente un carré de 3 nm de côté. Ils estiment également que les molécules simples ont une vitesse de diffusion moyenne de 30 nm/ms, alors que les macromolécules sont plus lentes avec une vitesse de diffusion moyenne estimée à 3 nm/ms. Par simplification, SimCell gère 1 déplacement des macromolécules tous les 10 pas de temps à la différence du mouvement des petites molécules qui est géré à chaque pas de temps.

Lors de la définition d'une espèce, on lui associe une couleur, or la palette disponible comprend 7 teintes et 10 nuances par teinte, soit 70 couleurs au total. De fait, SimCell est limité à 70 espèces. Si ce nombre peut déjà paraître conséquent, il est très vite atteint. Construire complètement la partie centrale du métabolisme constituée de la glycolyse et du cycle de Krebs requiert pas moins de 53 entités (20 métabolites, 21 enzymes, 12 métabolites intermédiaires).

De plus, lorsque le nombre d'intervenants est important, la lisibilité de l'automate cellulaire devient compliquée tout comme l'exploitation à l'aide des graphes.

De par ses limitations, et de par son approche basée sur les niveaux les plus bas de l'organisation cellulaire, en étudiant les molécules individuellement dans un espace, SimCell se prête à la simulation de réseaux biochimiques simples dans un environnement réduit. Il donne alors de bons résultats pour étudier le mouvement brownien et mettre en évidence le rôle de la chance dans le bon déroulement et la régulation des processus cellulaires, une réaction ne peut avoir lieu si tous ces intervenants ne se rencontrent au sein de l'espace (Wishart *et al.*, 2004 [68]).

B. V-CELL

V-Cell est disponible en ligne ^{2,3}(V-Cell, [69] ; [70]).

V-Cell se veut un outil de modélisation des processus cellulaires à destination de la recherche en biologie cellulaire. Les auteurs mettent pour cela en avant la capacité à gérer

² <http://vcell.org/>

³ <http://www.nrcam.uchc.edu/>

des modèles plus ou moins complexes et de natures différentes (mathématiques ou empiriques) exploitables à travers une unique interface. V-Cell est développé par le CCAM (Center for Cell Analysis and Modeling) depuis 1997 et repose sur les travaux de Schaff *et al.* (Schaff *et al.*, 1997 [71]). V-Cell est comparable à SimCell sur de nombreux aspects, mais il va beaucoup plus loin. Là où SimCell est contraint à 70 types d'entité, il n'est pas fait explicitement mention d'une limite pour V-Cell et l'espace géré n'est pas uniquement bidimensionnel, mais tridimensionnel, ou plus précisément un volume est géré par superposition de couches bidimensionnelles. Ceci nécessite des interactions entre les couches que V-Cell peut prendre en charge. De plus l'espace est structuré de façon beaucoup plus détaillée que SimCell permettant de prendre en compte la topographie et la géométrie des cellules telles qu'on peut les obtenir par des techniques d'imagerie. V-Cell permet d'ailleurs d'importer des données issues de techniques d'imagerie telles que l'IRM, la microscopie à contraste de phase, etc... À la différence de SimCell, V-Cell représente donc au minimum une cellule.

La puissance de V-Cell se paie par une moindre ergonomie et nécessite de l'utilisateur plus de compétences et plus de temps pour le paramétrage. En contrepartie, il n'y a pas de limite à la définition des caractéristiques des espèces modélisées ni aux interactions prises en compte. Tout peut être défini par l'utilisateur à travers l'interface graphique (Figure III.14).

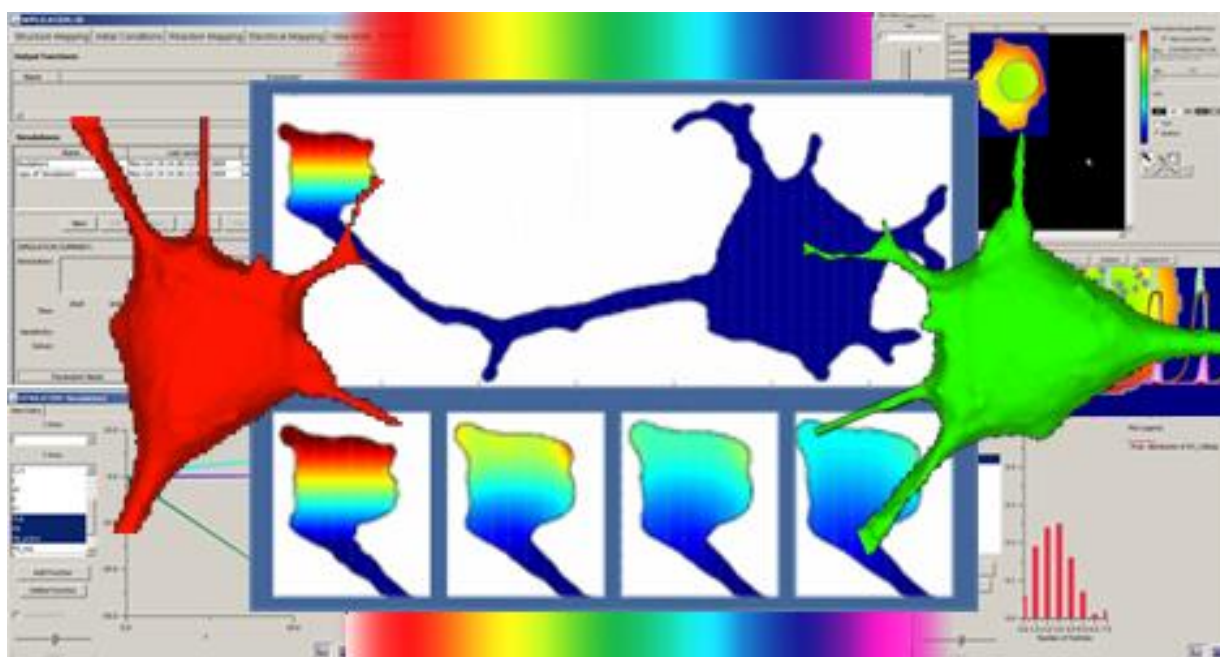


Figure III.14 : Exemple de l'interface de V-Cell dans la modélisation de la transmission d'un influx nerveux au niveau de l'axone d'un neurone isolé [72].

Par son approche, V-Cell permet d'étudier finement les processus à l'œuvre dans une cellule mais avec une moindre précision que SimCell, V-Cell gère les interactions moléculaires par une approche moins élémentaires, sans regarder au cas par cas mais par une approche statistique. Il se révèle donc moins adapté pour le suivi des détails des processus élémentaires comme par exemple la diffusion mais il se révèle plus adapté à la compréhension du fonctionnement d'une cellule, comme par exemple l'organisation cellulaire (Slepchenko *et al.*, 2007 [73]), la dynamique du cytosquelette (Ditlev *et al.*, 2009

[74]) ou l'origine biochimique de phénomènes électro-physiologiques des neurones (Brown *et al.*, 2010 [75]).

C. SNOOPY

Snoopy est disponible à en ligne ⁴ (Snoopy, [76]).

Il s'agit d'une approche intéressante pour modéliser les réseaux métaboliques qui passe par l'utilisation de réseaux de Petri. Les réseaux de Petri ont été formalisés par Carl Adam Petri dans sa thèse en 1962 (Petri, 1962 [77]). Dans un réseau de Petri on trouve des transitions et des places. Les places ne sont « que des espaces de stockage » entre les transitions. Les places et transitions sont connectées par des arcs.

Dans un tel modèle utilisé pour modéliser les processus biochimiques, ce qui est paramétré ce ne sont pas les espèces chimiques en présence, mais les interactions qui existent entre elles et qui seront représentées par les transitions et les arcs.

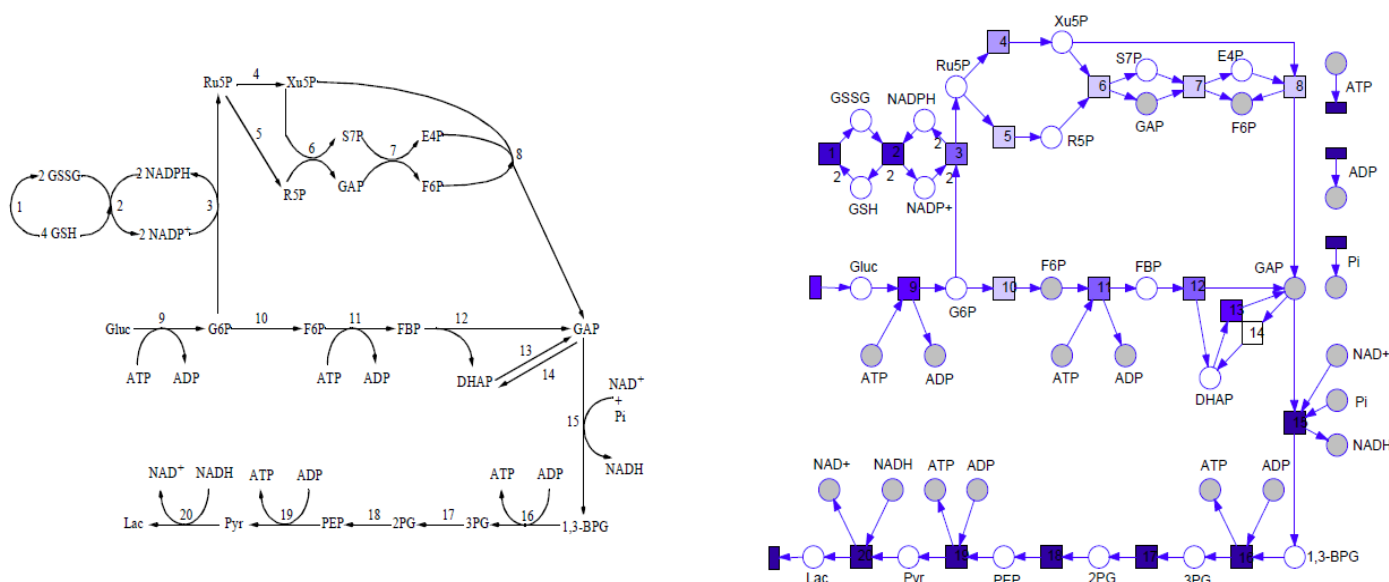


Figure III.15 : Exemple de voie métabolique représentée sous forme d'un réseau de Petri, la glycolyse (Embden-Meyerhof et pentoses-phosphate) [78].

Les réactifs et produits des différentes réactions qui peuvent être envisagés n'apparaissent pas explicitement dans ce genre de modèle, ils sont représentés par les jetons qui seront disposés sur les places. Si une interaction est réversible, cela nécessite la conception de deux transitions, une pour chaque sens de la réaction. Chaque transition peut être paramétrée indépendamment l'une de l'autre. Ainsi, si une enzyme est impliquée dans plusieurs réactions, chaque réaction sera définie par une transition avec son propre jeu de paramètres, si bien que ce genre de modèle offre une grande finesse dans la modélisation de réseaux métaboliques (Figure III.15).

⁴ <http://www-dssz.informatik.tu-cottbus.de/software/snoopy.html>

Le déroulement d'une telle simulation se passe en injectant des jetons dans le réseau. Les jetons passent alors de place en place selon les transitions. Une transition peut nécessiter des seuils pour se réaliser, et tant qu'il n'y aura pas assez de jetons en entrée elle n'aura pas lieu.

L'approche des réseaux de Petri pour modéliser les réseaux métaboliques a produit des résultats intéressants dans la régulation du métabolisme chez *E. coli* (Marwan *et al.*, 2012 [79]).

III.3.2. ADAPTATION

L'adaptation consiste pour un organisme à apporter une réponse appropriée à une stimulation quelconque, que ce soit un changement dans son environnement, un dysfonctionnement interne, ou un message émis par un congénère. En ce sens, les phénomènes observés sont très rapides en comparaison à la vie de l'individu.

Observer la réponse de chaque individu d'une population donnera un aperçu de l'adaptation de la population. Ainsi les modèles dont l'objet est l'adaptation s'intéressent à l'équilibre et la régulation de populations par les interactions qui existent entre leurs individus sous la contrainte de l'environnement. Ils s'appuient donc sur des modèles individu-centrés pour produire des observables à échelle de la population.

Les modèles cités précédemment répondant à la classification « Différenciation-Migration-Mort » appartiennent tous à cette catégorie. En effet de tels modèles couvrent la vie de l'individu et s'intéressent au détail du devenir de chacun selon les interactions qu'ils entretiennent entre eux et avec leur environnement, en fonction de leur déplacement et de leur arrivée dans un nouvel environnement.

Nous avons choisi de présenter deux exemples de tels modèles, BacSim et INDISIM qui partagent un même principe mais reposent sur des hypothèses de modélisation différentes.

A. BACSIM

BacSim a été développé en 1998 par Jan-Ulrich Kreft (Kreft *et al.*, 1998 [80], Figure III.16) sur la base de GECKO, un système multi-agents utilisé pour l'étude de la dynamique des écosystèmes écrit par Ginger Booth (Booth, 1997 [81]). Il est disponible en ligne ⁵ (BacSim, [82]).

Ce modèle a pour objectif de modéliser la croissance bactérienne et le développement de biofilms, il considère pour cela un environnement donné dans lequel des bactéries assurent leur survie et se multiplient en utilisant les nutriments disponibles, du glucose. Si à un moment donné la cellule n'obtient plus suffisamment de ressources pour se maintenir, elle mourra. Le glucose disponible dans l'environnement limite la croissance et la vitesse de multiplication des cellules.

L'objectif avec BacSim était d'intégrer les processus subcellulaires au modèle global de la cellule. Chaque cellule est modélisée par un agent autonome doté de ses propres

⁵ http://www.theobio.uni-bonn.de/people/jan_kreft/bacsim.html

paramètres. Les auteurs considèrent que les cellules sont sphériques et que leur capacité d'absorption du glucose est liée à la surface de leur membrane, ainsi plus une cellule est grosse, plus elle peut absorber de nutriments à chaque pas de temps de la simulation. Pour cela, ils considèrent que la génétique de la cellule détermine le nombre de transporteurs qui peuvent prendre place par unité de surface de la membrane cellulaire.

Le glucose absorbé permet à la cellule de grossir. Plus une cellule absorbe de nutriments, plus elle grossit, plus sa surface se développe, et plus elle peut absorber de nutriments, etc... Inversement, un déficit en glucose se traduira par une réduction du volume cellulaire.

Lorsqu'une cellule atteint un volume maximal elle se divise et donne naissance à deux nouvelles cellules dont le volume est la moitié du volume de la cellule mère. Si au contraire le manque de nourriture amène la cellule à un volume minimum elle meurt. Dans ce cas, son contenu retourne à l'environnement.

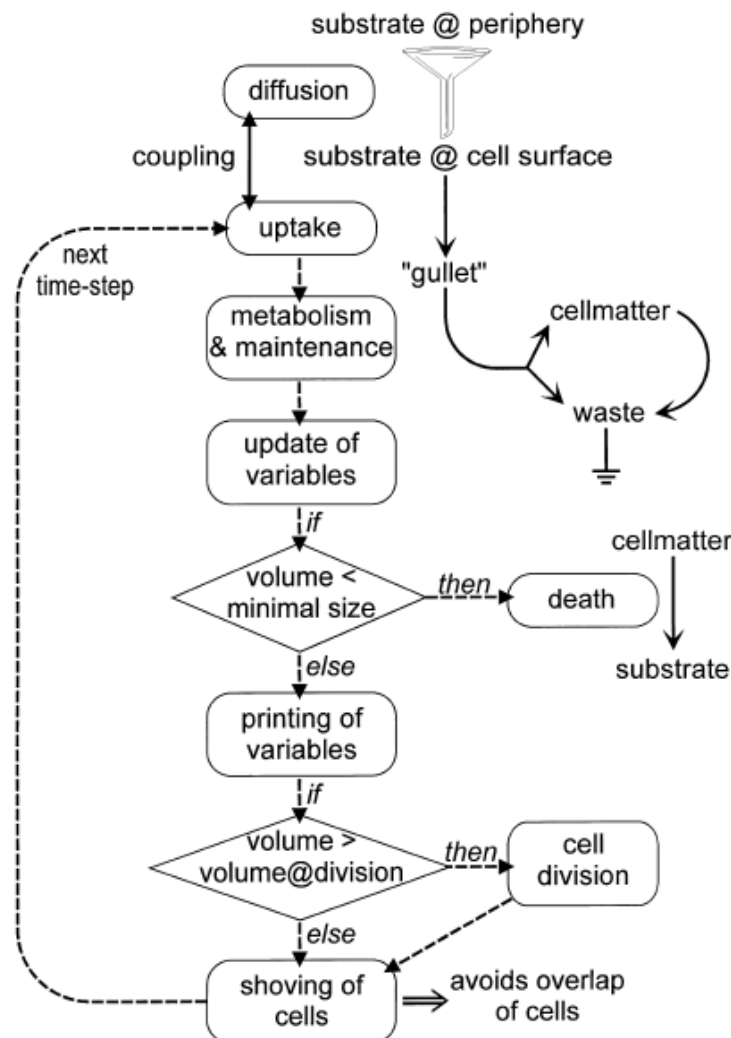


Figure III.16 : L'algorithme de BacSim tels qu'il est présenté par ses auteurs [80].

Le glucose se diffuse dans l'environnement pour parvenir aux cellules, ainsi, si une cellule absorbe le glucose qui l'entoure plus vite qu'il ne diffuse, c'est la quantité de nutriments disponible à chaque cellule qui est l'élément limitant de la croissance.

Dans le détail, tous les nutriments absorbés par une cellule ne participent pas à son augmentation de volume, une partie est ponctionnée par ce que les auteurs appellent la maintenance qui définit le maintien en bon état de marche du matériel cellulaire. Partant du principe que plus une cellule est volumineuse, plus elle a de matière à entretenir, la quantité de glucose consommée dévolue à la maintenance est une proportion fixe de ce qui est consommé.

La quantité de glucose que peut absorber une cellule par unité de surface membranaire et par pas de temps, le volume minimum, le volume maximum et la proportion de nutriments nécessaire à la maintenance sont des paramètres qui appartiennent à une liste par défaut. Cette dernière représente l'espèce, le patrimoine génétique commun à chaque individu. Plusieurs jeux de paramètres ont été testés, représentant respectivement *Escherichia coli* ou *Bacillus subtilis*.

Des mutations peuvent survenir lors de la formation des nouvelles cellules. Les paramètres que reçoit alors une cellule mutante sont modifiés suivant une loi gaussienne.

Hormis les quelques mutations, toutes les cellules présentent les mêmes caractéristiques, mais comme chaque cellule dispose de son environnement propre et donc d'une quantité de glucose propre, chaque cellule se développe à son propre rythme.

C'est un automate cellulaire qui gère la diffusion des nutriments dans l'environnement. Les cellules n'y apparaissent pas explicitement mais sous la forme de sites de disparition du glucose. Les cellules n'ont pas de connaissance de l'espace ou de leur position.

BacSim a par la suite été utilisé en 2005 (Dens *et al.*, 2005 [83] ; [84]) pour tester l'effet d'un changement de température pendant la croissance cellulaire en utilisant les paramètres pour *E. coli*. L'hypothèse prise est que la température influence le volume de la cellule, si la température baisse le liquide se contracte, la cellule réduit son volume et donc sa surface, et par conséquent ce sont ses capacités d'absorption du glucose qui sont diminuées. Inversement, si la température augmente, la cellule se dilate, sa surface augmente et sa capacité de consommation progresse dans le même temps. Ceci se traduit par un changement de la vitesse de croissance lorsque la température de l'environnement est changée.

B. INDISIM

INDISIM a été développé par Marta Ginovart en 2002 (Ginovart *et al.*, 2002 [85], Figure III.17). Selon la définition de ses auteurs INDISIM partage la même philosophie que d'autres modèles individu-centrés comme BacSim, s'attachant à la description de cellules bactériennes en culture, en gèrent les propriétés des bactéries, leur consommation, leur métabolisme, la division et la mort des cellules. La différence de leur approche par rapport à BacSim se situe dans la gestion de l'espace, là où BacSim utilise un automate cellulaire pour gérer les nutriments et un espace continu pour gérer la population bactérienne, INDISIM se

propose d'utiliser un seul espace discret pour gérer les nutriments et les cellules. Avec INDISIM, les auteurs proposent également de gérer le mouvement des cellules. Chaque maille de l'espace peut accueillir jusqu'à quatre cellules.

Comme dans BacSim, INDISIM propose de gérer l'absorption de nutriments par les cellules en considérant le nombre de nutriments accessibles, ceux-ci diffusent dans l'espace. La quantité de nutriments qui peut être absorbée par cellule et par pas de temps est liée à l'unité de biomasse de la cellule. Les auteurs considèrent ainsi que le taux de transporteurs et d'enzymes permettant l'assimilation des nutriments est une proportion constante de la matière cellulaire définie par la génétique. Une partie de ce qui est consommé sert à assurer la maintenance des fonctions cellulaires. Cette partie est également proportionnelle à la masse cellulaire, le reste sert à augmenter la biomasse. Ainsi, BacSim et INDISIM partagent un principe identique, mais dans le second, le terme de biomasse remplace la notion de volume employée dans le premier pour désigner le matériel cellulaire.

Une cellule qui ne peut assurer sa maintenance voit sa biomasse diminuer en conséquence. Si la biomasse atteint une valeur minimale critique, la cellule meurt, si au contraire elle augmente jusqu'à une valeur maximale que les auteurs appellent masse spécifique, la cellule se divise et donne naissance à deux cellules filles, chacune représentant environ la moitié de la masse de la cellule mère.

Autre différence importante par rapport à BacSim, la division n'est ici pas considérée comme un processus instantané mais a une durée définie par les propriétés de l'espèce de bactérie considérée, identiques pour toutes les bactéries de cette espèce dans les conditions de l'expérience. Ce processus est irréversible, une cellule qui commence à se diviser termine sa division même si l'environnement ne présente plus de nutriments.

Dernier point notable, les nutriments présents dans l'environnement ne sont pas forcément accessibles aux cellules, dans ce cas la cellule doit produire des enzymes extracellulaires qui vont transformer les nutriments de l'environnement en une forme assimilable par les cellules. Il existe pour chaque cellule une capacité de production de ces enzymes extracellulaires qui est liée à la biomasse cellulaire, le taux et le rendement de conversion des nutriments par ses enzymes.

Comme dans BacSim, c'est un jeu de paramètres qui représente le génome, mais leur nombre est plus important dans INDISIM présentant la capacité de consommation maximale par unité de masse, l'efficacité du métabolisme (la proportion de nutriments qui sont utilisés pour la maintenance ou convertis en matière cellulaire), les besoins en nutriments pour assurer la maintenance d'une unité de biomasse, la masse spécifique, la masse minimale et la durée de la division. Éventuellement, se rajoutent les paramètres concernant les enzymes extracellulaires, leur taux de production en fonction de la masse cellulaire, leur taux de conversion des nutriments en produits assimilables, et le rendement de cette conversion.

INDISIM a été utilisé pour modéliser la phase de latence dans la croissance bactérienne (Prats *et al.*, 2006 [86]), cette phase qui survient avant la phase de croissance exponentielle peut avoir deux causes. D'une part, il peut s'agir de cellules âgées/abimées qui auront besoin de se « restaurer » avant de pouvoir reprendre une croissance normale. D'autre part il peut s'agir de cellules qui sont déplacées d'un environnement à un autre. La latence est alors due à

l'adaptation des cellules à leur nouvel environnement. Les deux causes ont été étudiées à l'aide de ce modèle. Pour évaluer la première, ils ont simulé l'inoculation du milieu par une population de cellules dont la biomasse était faible et hétérogène, la croissance exponentielle ne survient qu'après que les cellules initiale aient grossi suffisamment pour reprendre leur croissance. Pour le changement de milieu, ils ont considéré que les cellules utilisées pour l'inoculation du milieu avaient une biomasse moyenne et homogène, et qu'elles disposaient d'enzymes pour utiliser un type de nutriments, mais pas pour l'autre, le changement de ressource se traduit par un temps pendant lequel les cellules réduisent leur biomasse pour produire les enzymes permettant l'utilisation du nouveau nutriment. La croissance exponentielle réapparaît quand les cellules ont rétabli cette capacité à extraire des nutriments de leur environnement et récupéré une masse suffisante.

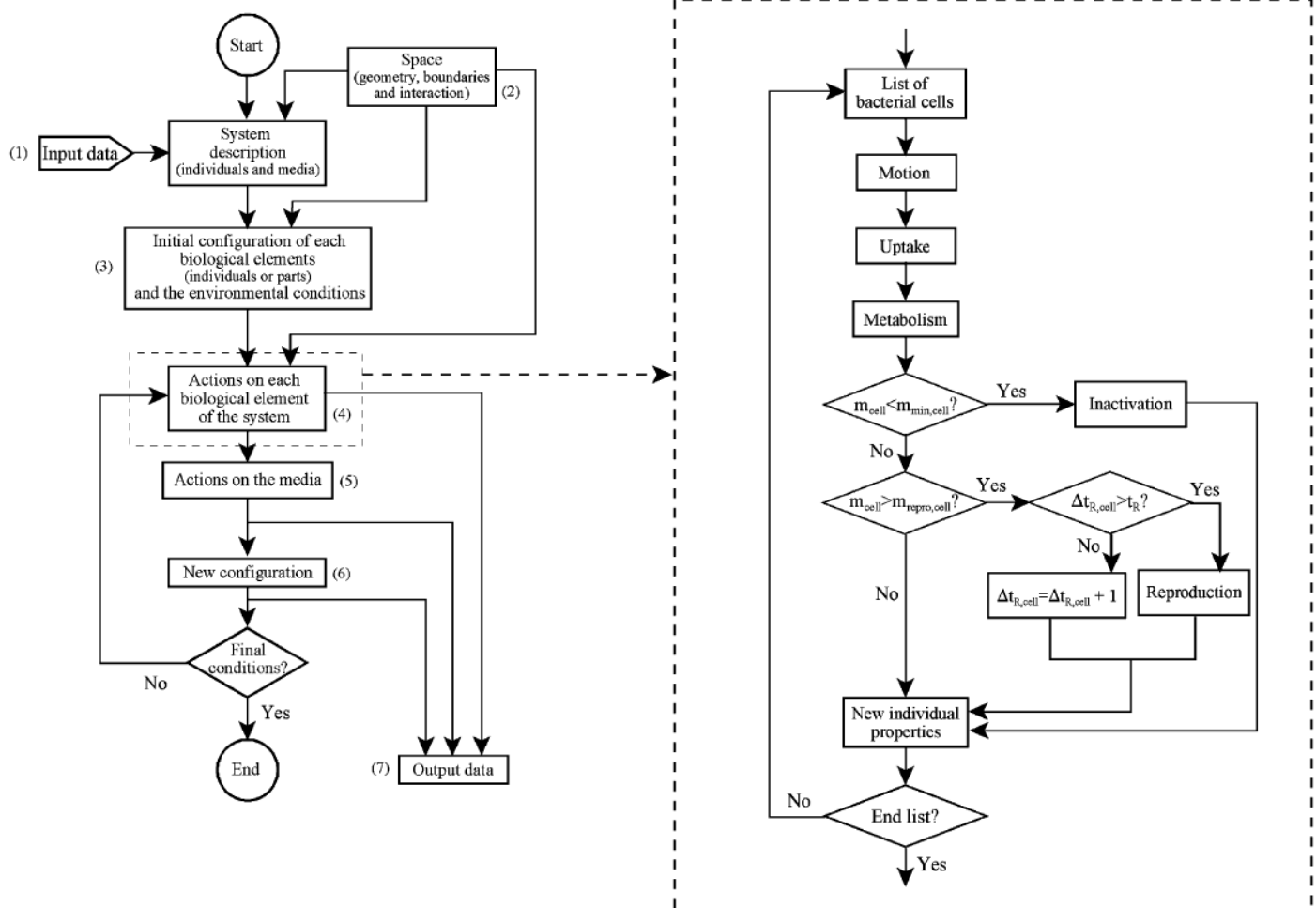


Figure III.17 : L'algorithme de INDISIM tel qu'il est présenté par ses auteurs [86].

INDISIM a servi de base à plusieurs travaux pour lesquels diverses adaptations ont été réalisées en fonction de la problématique adressée. INDISIM-SOM (Ginovart *et al.*, 2005 [87] ; Gras et Ginovart, 2006 [88]) est une extension de INDISIM qui se veut spécialisée sur l'étude de l'activité microbienne dans les sols (SOM = Soils Organic Matter) en permettant de manipuler deux classes de bactéries avec des propriétés distinctes, des microorganismes décomposeurs et des microorganismes nitrifiants impliqués respectivement dans le cycle du carbone et dans le cycle de l'azote. Pour chaque type de bactéries il existe un jeu de

paramètres et l'environnement présente 9 types de nutriments. La compétition entre les bactéries pour certaines ressources et les interrelations qui existent entre les fonctionnements des différents types d'organismes produisent le comportement global du système. Une version d'INDISIM est spécialisée dans la modélisation de levures INDISIM-YEAST (Ginovart *et al.*, 2007 [89]), les levures ne se divisent pas de façon symétrique mais par bourgeonnement : une petite cellule, le bourgeon se forme à la surface d'une cellule mère qui reste identifiable, contrairement avec ce qui se passe avec les bactéries où la cellule mère disparaît remplacée par deux cellules filles. Une levure ne peut bourgeonner qu'un nombre limité de fois. INDISIM-YEAST est modifié pour prendre en compte ces spécificités du fonctionnement des levures, le nombre de divisions que peut produire une cellule est limité et elle en conserve « la mémoire », et la division n'est pas symétrique en termes de biomasse. En 2014 (Portell *et al.*, 2014 [90]), INDISIM-YEAST a été étendu à un système tridimensionnel sous le nom d'INDISIM-Saccha.

III.3.3. EVOLUTION

L'évolution, au sens darwinien du terme, c'est la transformation de populations d'organismes sur des périodes longues en réponse à la modification de leur environnement.

Dans les systèmes informatiques intelligents, capable d'apprendre ou d'évoluer, l'évolution est décrite comme « adaptation collective ».

Les modèles qui s'y intéressent cherchent à reproduire cette modification des individus au fil des générations sous la pression d'un moyen de sélection lié à un changement de l'environnement ou la compétition entre les individus. Ils s'appuient sur des modèles individu-centrés et produisent des observables à échelle de la population.

Puisque c'est une évolution que l'on cherche à observer, les modifications entre les individus par génération sont « exagérées » et « amplifiées », c'est-à-dire que les modifications modélisées sont plus importantes et plus fréquentes que ce qui s'observe dans la nature, et la multiplication des individus est également accélérée, la vie propre des individus ayant peu d'importance. L'échelle de temps considérée est la génération ou une échelle de temps très longue par rapport à la vie de l'individu, par exemple pour des bactéries, elle pourrait être de l'ordre de l'année, la vie de l'individu étant de quelques dizaines de minutes.

Chaque individu possède son propre patrimoine. Ce patrimoine porte un ou plusieurs caractères. Plusieurs individus peuvent partager le même patrimoine ou les mêmes caractères. C'est sur ces patrimoines ou caractères que vont s'exercer les changements, et c'est par eux que va pouvoir être mis en place un processus de sélection.

L'évolution étant un processus lent dans la nature, difficile à appréhender, et pour lequel il n'est pas facile de proposer des expériences ou des observations en laboratoire ou dans un environnement quelconque, elle a été modélisée par des approches très différentes. Nous avons choisi d'en présenter quatre : Tierra, Evolved Virtual Creatures, Avida, Aevol.

A. TIERRA

Tierra est disponible en ligne ⁶ (Tierra, [91]).

Tierra a été développé en 1991 par Thomas Ray (Ray, 1991 [92]). C'est la première plateforme dédiée à l'étude de l'évolution *in silico*. Il part du constat que si l'autoréplication mise en avant par Von Neumann (Von Neumann, 1966 [25]) est importante pour le vivant, elle ne permet pas de rendre compte de la diversité du vivant. L'idée était alors de développer une structure au sein de laquelle l'autoréplication puisse se réaliser, mais sans que la structure ou le support de l'objet qui se réplique ne soit une contrainte, c'est-à-dire introduire la possibilité pour un objet de s'autoreproduire mais avec des modifications. Tierra a été mis à jour pour la dernière fois en 2004.

Dans ce modèle, les objets désignent des organismes. Un organisme est constitué de trois éléments, une séquence chromosomique qui contient toute l'information pour le reproduire, un « interpréteur » qui lit le « chromosome » et le réplique, et un « scanner » qui repère un espace mémoire assez grand pour recevoir une copie de l'organisme. Chaque élément est constitué d'une séquence « double brin » de triplets de bits. L'appariement entre les brins est contraint par la séquence de bits. Par exemple le triplet 101 s'apparie avec le triplet 010.

Le chromosome peut subir des mutations. Les mutations peuvent affecter aléatoirement n'importe quel individu présent à tout instant. Parmi les possibilités prises en compte pour les mutations on va retrouver le changement aléatoire d'un bit, ou la permutation de deux bits consécutifs ou appariés, ou encore la recombinaison entre deux séquences occupant des cases mémoires proches

Une mutation sur un bit d'un triplet sera aléatoirement reportée sur le brin complémentaire ou réparée en se basant sur le brin complémentaire. Par exemple, si le triplet 101 apparié avec le triplet 010 subit une mutation sur son troisième bit pour devenir le triplet 100. Soit il sera réparé pour être rapparié correctement au triplet 010 qui est en face de lui et redeviendra le triplet 101, soit le triplet 010 sera modifié en 011 pour être correctement apparié au nouveau triplet 100.

La mort existe dans ce modèle et elle peut frapper aléatoirement un petit nombre d'individus à n'importe quel moment. Un organisme qui apparaît perdure tant que la mort ne le frappe pas, il n'y a pas de modèle de durée de vie des organismes. Comme la mort est un processus aléatoire, certains individus peuvent avoir une durée de vie équivalente à la durée de la simulation alors que d'autres seront éliminés aussitôt générés. Un organisme qui meurt libère simplement les ressources qu'il occupait.

La séquence du chromosome décrit trois gènes. L'interprétation du premier permet le calcul de la taille totale du chromosome. Le second gène code l'interpréteur qui va permettre l'expression des deux autres gènes et la recopie du chromosome dans un nouvel espace mémoire, et l'expression du troisième gène crée un « scanner » qui cherche un espace mémoire suffisant pour accueillir le nouvel individu.

⁶ <http://life.ou.edu/tierra/>

Les organismes sont en concurrence pour le temps de calcul processeur qui représente l'énergie disponible dans le système, et pour la mémoire qui correspond aux ressources de l'environnement. Les organismes évoluent au cours de la simulation en améliorant l'efficacité de leur interpréteur pour qu'il nécessite moins de temps de calcul, en réduisant l'espace mémoire qu'ils occupent afin de se multiplier plus vite et d'être plus nombreux, ou au contraire en augmentant leur taille pour occuper la mémoire et bloquer la prolifération d'organismes concurrents.

Tierra a été utilisé pour réaliser des simulations sur grille de calcul et sur supercalculateur donnant à chaque fois naissance à des écosystèmes uniques. Il a été observé une mutation fréquente qui donne naissance à des « virus » dépourvus d'interpréteur qui parasitent un individu voisin en détournant son interpréteur pour se répliquer. Des équilibres s'instaurent spontanément dans les populations modélisées et on peut par exemple observer l'exclusion de certains types d'individus de certaines zones, ou au contraire observer une cohabitation harmonieuse des populations en présence comme par exemple le rapport entre les parasites et leurs hôtes. Ce modèle permet de mettre en évidence le rôle du parasitisme et l'importance de la « chance » dans la diversification des formes du vivant.

B. EVOLVED VIRTUAL CREATURE

Evolved Virtual Creatures a été créé en 1994 par Karl Sims (Karl Sims, 1994 [93] ; [94]). Le modèle semble avoir disparu d'internet, il reste cependant en ligne une implémentation réalisée par Thomas Ray en 2001 ⁷ (Virtual Life, [95]).

L'objectif affiché est de tester les capacités d'un programme à résoudre un problème simple, aller d'un point de départ fixe à une cible donnée. Le programme génère des formes aléatoirement, articulées et mobiles. On compare la stratégie choisie pour effectuer le déplacement de chaque structure ainsi élaborée et on sélectionne les plus efficaces. L'idée est ainsi de découvrir comment les stratégies de déplacement des organismes vivants se sont mises en place dans la nature.

Pour cela, ce modèle produit une créature en assemblant un nombre aléatoire de volumes réguliers à des positions aléatoires et selon des angles aléatoires. Pour une créature générée, on associe un « chromosome » qui porte l'information du nombre de polygones, les points et angles d'attache, et les degrés de mouvements à chaque articulation. La créature ainsi générée est placée sur une surface bidimensionnelle dans un champ de force mimant l'action de la gravité et elle doit rejoindre une « source de nourriture ». La créature doit trouver comment se mouvoir dans l'espace en actionnant ses différents volumes. Chaque créature porte un « système nerveux » qui coordonne ses mouvements.

Thomas Ray a enrichi le modèle en incluant un champ de force mimant un environnement aquatique, et a ainsi testé la capacité de ce modèle à développer des stratégies de natation (Ray, 2001 [96]).

À partir d'une créature, l'évolution sera modélisée en apportant des modifications aléatoires au chromosome de la créature, ce qui changera le nombre d'éléments, leur position et les

⁷ <http://life.ou.edu/VirtualLife/>

angles de liaison qui la compose, de sorte que la nouvelle créature générée se déplacera différemment de la créature précédente. Si elle atteint plus rapidement l'endroit désigné que la créature « mère » on la conserve, sinon, on revient à la créature « mère » et on procède à de nouveaux changements aléatoires sur son chromosome.

Ce modèle d'évolution est assez particulier puisque il ne modélise qu'un ou deux individus à la fois, il n'y donc pas de compétition entre les individus pour la sélection, mais seulement une comparaison de la performance des uns et des autres. La sélection se fait entre les générations selon un critère unique, l'efficacité à rejoindre une source de nourriture. Il n'y a pas d'interactions.

Ce modèle a permis l'émergence de plusieurs types de déplacements comme la marche, la reptation, l'utilisation d'une ou plusieurs nageoires pour nager, mais aussi des modes de déplacement inconnus ou inutilisés par les créatures de notre Terre, comme par exemple la formation d'hélice pour nager, ou un déplacement terrestre utilisant un pied unique pour enchaîner les culbutes.

C. AVIDA/AVIDA-ED

Avida est un projet débuté en 1994 par Chris Adami et C. Titus Brown (Adami et Brown, 1994 [97]). Avida est disponible en ligne ⁸ (Avida, [98]).

Avida s'inspire de Tierra mais pousse le concept plus loin. Là où les organismes virtuels de Tierra ne sont représentés que par trois gènes et que tous dépendent du même processeur et du même espace mémoire pour lesquels ils entrent en concurrence, dans Avida chaque organisme virtuel est une machine virtuelle complète disposant de son espace mémoire et de son « processeur ». Le chromosome de ces organismes virtuels est beaucoup plus élaboré avec des suites complètes d'instructions dans un langage compris par le « processeur » de l'organisme. Les organismes virtuels sont disposés sur un automate cellulaire, et l'application des fonctions de leur chromosome permet la mise à jour de l'automate.

Les règles d'interaction entre un organisme et son voisinage permettent des échanges de matériel chromosomique entre les individus, ce qui représente la reproduction sexuée. S'il n'y a pas d'interaction, l'organisme virtuel se duplique simplement sur une cellule proche dans l'automate cellulaire.

Comme chaque organisme dispose de son propre « processeur » et de son propre espace mémoire, ils peuvent chacun fonctionner à une fréquence différente, si bien que la compétition entre les organismes virtuels naît de la compétition pour la conquête de l'automate cellulaire, les organismes les plus efficaces étant les plus rapides à se reproduire et donc à coloniser l'automate. L'efficacité est liée à la fréquence de chaque « processeur », à la taille et à l'efficacité des chromosomes des organismes virtuels.

Pour comprendre ces deux derniers aspects, il faut envisager que dans Avida, les chromosomes des organismes virtuels présentent une suite d'instructions appartenant à un ensemble de 26 fonctions. Les mutations qui peuvent survenir sont des réécritures de ces

⁸ <http://avida.devosoft.org/>

fonctions, changeant la longueur de code qui les décrit, insérant des instructions inutiles, etc... Les instructions sont exécutées séquentiellement, ou selon un ordre donné par l'une des instructions. Ainsi, l'exécution du chromosome d'un individu peut sauter des instructions qu'il porte ou au contraire en faire exécuter certaines à plusieurs reprises.

Avida a été repris en 2003 par Richard Lenski (Lenski *et al.*, 2003 [99]) qui y a vu une démonstration *in silico* de la théorie de l'évolution de Darwin.

L'évolution est un processus trop lent pour être observé en laboratoire et surtout dans le cadre d'expériences dans un cursus scolaire. Un monde virtuel est plus facile à manipuler et à exploiter, permettant une compréhension facilitée de ce processus difficile à appréhender dans la nature ou en laboratoire. À partir de ce constat et sur la base d'Avida, Robert Pennock se propose de répondre à un besoin pédagogique pluridisciplinaire, traitant d'aspects biologiques, écologiques, et génétiques pour réaliser l'évolution au sein d'un mode virtuel et il développe à cette fin AvidaED (Figure III.18). Avida ED a été décrit en 2006 (Pennock, 2006 [100]) avant d'être mis à disposition en 2007 (Pennock, 2007 [101]) en ligne⁹ (Pennock, [102]).

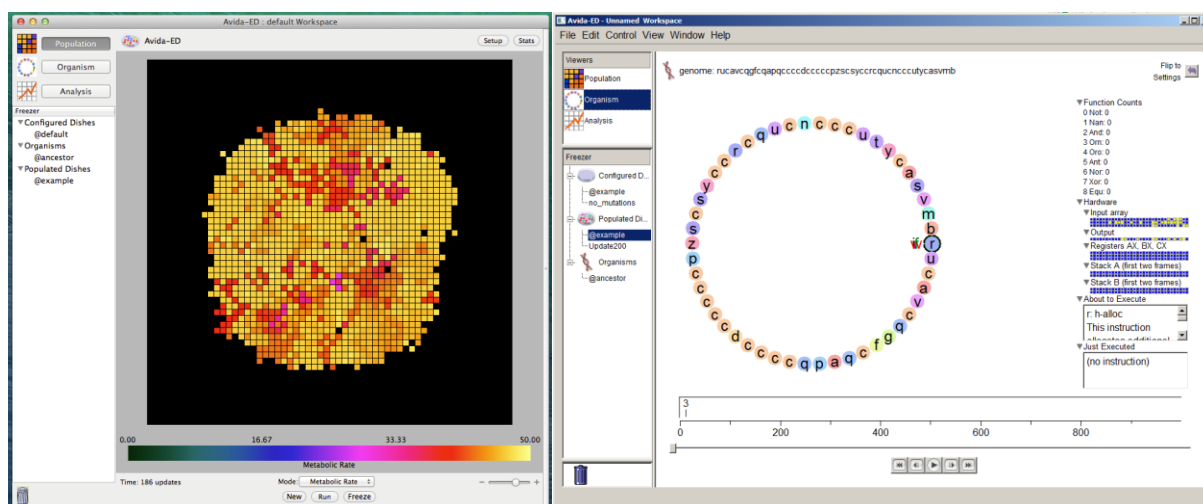


Figure III.18 : Captures d'écran d'AvidaED, à gauche l'interface de suivi de la population, à droite la fenêtre d'étude du chromosome d'un organisme virtuel.

L'environnement décrit par AvidaED représente une boîte de Petri, qui peut à tout instant être sauvegardée. Une nouvelle simulation peut être lancée à partir d'une sauvegarde pour voir les différences qui peuvent apparaître entre deux instances de simulation. Les organismes virtuels représentent des microorganismes qui peuvent être étudiés individuellement. L'évolution se fait par sélection selon l'efficacité des microorganismes, cette dernière étant évaluée en fonction de l'efficacité de leur chromosome et de sa taille.

D. AEVOL

Aevol est disponible en ligne¹⁰ (Aevol, [103]).

⁹ <http://avida-ed.msu.edu/>

¹⁰ <http://www.aevol.fr/>

Le but de ce modèle est de fournir un cadre expérimental pour tester l'influence de la structure des génomes dans le processus évolutif. Pour cela, sa configuration revêt une importance particulière et les opérateurs de mutation pris en considération seront plus variés et représenteront mieux ce qui existe dans le vivant.

Comme Tierra et Avida, Aevol modélise des organismes virtuels par leur génome représenté par un chromosome circulaire, constitué d'une suite de valeurs binaires. Ce chromosome est double brin et leurs séquences sont complémentaires, un 0 s'appariant à un 1. La différence introduite est qu'ici le génome est structuré en séquences codantes et séquences non-codantes.

Autre différence notable, l'expression du génome n'est pas l'élément sur lequel se fait la sélection, ou du moins, pas directement. En effet, dans Aevol, les auteurs ont modélisé le phénotype par simulation des processus de transcription et de traduction du génome que l'on retrouve dans le vivant. L'expression du génotype donne ainsi un phénotype qui sera comparé à celui des autres individus.

Les séquences codantes sont délimitées par un promoteur et par un terminateur. La transcription répond à la même règle d'appariement que la séquence chromosomique. La traduction qui intervient ensuite donne à chaque séquence transcrite une valeur, que les auteurs rapprochent des protéines.

C'est en comparant les valeurs de toutes les protéines de chaque individu, le phénotype, que les meilleurs vont être sélectionnés. C'est par l'environnement qu'on définit si un individu est meilleur qu'un autre. L'environnement est représenté par une fonction qui prend en argument les valeurs des protéines.

Alors que dans les modèles cités précédemment l'effectif de la population changeait dans le temps, Aevol travaille à population constante. La population est entièrement reconstituée à chaque pas de temps. Plus les individus sont adaptés, et plus la proportion d'individus partageant leur phénotype sera importante à la génération suivante.

Le principal intérêt d'Aevol réside dans sa gestion des opérateurs de mutation. Aevol est une version modifiée de l'algorithme RBF-Gene (Radial Basis Function) présenté en 2003 par Guillaume Beslon (Beslon *et al.*, 2003 [104], Lefort *et al.*, 2004 [105]) que les auteurs décrivent comme un algorithme génétique modifié. Il intègre aux chromosomes les séquences non-codantes. Ces dernières peuvent devenir des séquences codantes suite à une mutation, et inversement. Les opérateurs de mutation considérés sont de trois types, les opérateurs dits « locaux » qui ne concernent qu'un seul bit du chromosome, les opérateurs « larges » qui affectent plusieurs bits consécutifs du chromosome, et les opérateurs externes qui décrivent les échanges de matériel entre différents chromosomes.

À travers l'utilisation d'Aevol, l'importance de la structuration du génome dans la mise en place spontanée d'un niveau moyen de variabilité a été observée. Les lignées dans lesquelles la variabilité est trop importante disparaissent rapidement par instabilité, alors que les lignées au contraire trop stables sont supplantées dès lors qu'une autre lignée trouve une structure plus efficace (Knibbe *et al.*, 2006 [106] ; Batut *et al.*, 2013 [107]).

III.3.4. SYNTHÈSE DE CE TOUR D'HORIZON DES MODELES

Ce changement de perspective dans la classification des modèles existants permet de placer sur une même base tous les modèles, sans se limiter aux processus de migration, différenciation et mort. Cela met d'autant plus en évidence la difficulté de proposer un modèle multi-échelle capable de couvrir des échelles spatiales allant de l'ordre du nanomètre à celui du mètre ou plus, et des échelles de temps allant de la milliseconde à l'année (Figure III.19).

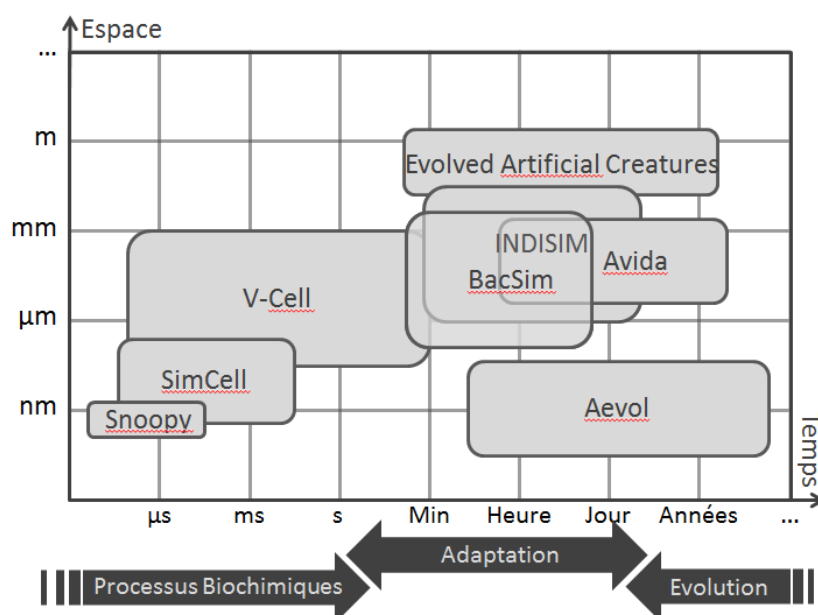


Figure III.19 : Échelles de temps et d'espace traitées par les quelques modèles évoqués.

Mais dans le même temps, cette classification et ce tour d'horizon des modèles existants confirment que notre objectif radiobiologique nécessite d'élaborer un modèle structuré de sorte qu'il permette la gestion des échelles impliquées dans les processus biochimiques pour contrôler le comportement des cellules en fonction des informations portées par leur patrimoine génétique mais soumises à l'influence de leur environnement et ce, sans perdre le devenir au sein de la population de la particularité de chaque individu pour modéliser l'évolution d'une population face à un stress radiatif, sa résilience face à ce stress ou sa transformation.

En effet, si la position centrale de la cellule dans le fonctionnement du vivant n'est pas remise en question, nous ne pourrions modéliser l'impact d'un stress radiatif sur les cellules sans gérer les processus biochimiques qui résulteront de l'interaction entre les rayonnements et la matière cellulaire, qu'elle ait lieu au niveau des protéines ou au niveau génomique, tant les différences dans le comportement de la cellule dépendront de la localisation des lésions.

Le détail des dégâts subis au niveau moléculaire, une fois observé à l'échelle cellulaire devra pouvoir être suivi dans la population. De fait, si une mutation survient, il n'y aura qu'en gérant des échelles de temps importantes qu'on pourra observer son devenir au sein de la population, et voir si la mutation favorisera une lignée de cellules, si elle dérégulera certains échanges, etc...

Toute la complexité du système vivant, de sa régulation, de ses échanges avec son environnement, de sa capacité de réaction aux changements ou aux perturbations qu'il subit s'exprime dans ces échelles de temps et d'espace, et nous ne parviendrons à la modéliser qu'à travers un modèle multi-échelle.

III.4. PLATEFORMES

En ayant parcouru les différents modèles de systèmes biologiques, nous avons constaté que pour tous il existe un environnement informatique ou logiciel qui permet leur exploitation. Observer tous ces environnements logiciels met en évidence deux types d'approches. Nous parlerons de plateformes pour ces environnements de modélisation. Certaines plateformes sont élaborées en parallèle à un modèle et lui sont dédiées, dans les exemples cités précédemment c'est le cas de SimCell, de BacSim, d'INDISIM, de Tierra, d'Evolved Virtual Creatures, d'Avida ou d'Aevol. Nous les qualifierons de plateformes dédiées.

Inversement, certaines plateformes sont prévues pour faire jouer différents modèles, c'est le cas de V-Cell, ou Snoopy par exemple. Ceci amène une notion de formalisme. Tant que les modèles sont décrits selon le formalisme requis, ils peuvent être simulés par ces plateformes que nous qualifions de génériques.

III.4.1. FORMALISMES

Les formalismes correspondent à des règles de définition et de description des modèles standardisées, permettant de les lire dans un même cadre avec une même méthodologie. Leur objectif est de permettre une uniformisation des règles d'écriture des modèles pour faciliter leur exploitation à travers des plateformes différentes, ou de permettre l'élaboration d'un modèle simplement sans avoir à se soucier du cadre de la simulation. L'idée sous-jacente étant de pouvoir focaliser sur le développement des modèles à proprement parler et non des outils qui *in fine* présenteront souvent des fonctions similaires.

Nous avons choisi d'en présenter brièvement deux : SBML et CellML.

A. SBML : SYSTEM BIOLOGY MARKUP LANGUAGE

SBML est un projet de langage de modélisation dérivé du standard XML qui a débuté en 2001 sous l'impulsion de Hiroaki Kitano (Kitano, 2001 [108]). L'objectif était d'uniformiser la façon de modéliser et d'écrire les modèles de processus biochimiques. La première version a été établie la même année (Hucka *et al.*, 2001 [109]) et publiée en 2003 (Hucka *et al.*, 2003 [110]). La dernière version (SBML Level 3 Version 1 Core) est disponible depuis 2010 (Hucka *et al.*, 2010 [111]). SBML est supporté par le System Biology Workbench (SBW).

Parmi les logiciels qui permettent d'éditer des modèles au format SBML, nous citerons CellDesigner (CellDesigner, [112]), V-Cell et Snoopy qui ont été décrits précédemment. La dernière version de ce projet date de l'été 2012.

B. CELLML :

CellML est un autre projet de langage qui dérive également du standard XML. Il est supporté par l'Institut de Bio-Ingénierie de l'Université d'Auckland. Les informations relatives à ce projet sont disponibles en ligne ¹¹ (CellML, [113]).

Le projet a commencé en 2000 avec les travaux de Waren Hedley et Melanie Nelson (Hedley *et al.*, 2000 [114]) autour du projet Physiome, destiné à la modélisation du fonctionnement des organes depuis le niveau cellulaire. La version 1.2 de CellML est en cours de développement, la dernière mise à jour ayant eu lieu au cours de l'été 2010.

Développer et exploiter des modèles selon ce formalisme peut se faire par le logiciel OpenCell (OpenCell, [115]), mais aussi par V-Cell qui permet à travers son interface d'élaborer et de simuler de tels modèles.

III.4.2. PLATEFORMES DEDIEES

Les exemples cités avant dans cet état de l'art sont pour la plupart des plateformes dédiées. Rappelons celles que nous avons vu qui se rattachent à cette catégorie, il s'agit de SimCell, de BacSim, d'INDISIM, de Tierra, d'Evolved Virtual Creatures, d'Avida et d'Aeol.

Il en existe d'autres, chacune conçue pour répondre à un besoin de modélisation particulier. Citons par exemple M-Cell (M-Cell, [116]), plateforme dédiée à la modélisation de processus biochimiques dont la particularité est d'être conçue pour tourner sur des grilles de calcul. Le modèle gère un volume qui peut être représenté par un cube dont le côté vaut 1 000 000. Sur une grille de calcul, chaque nœud exploite un volume de dimension similaire. C'est l'entité modélisée la plus petite qui représente l'unité de grandeur du volume (UV) et sa taille selon les trois axes de l'espace constitue respectivement l'échelle dans chacune des dimensions du volume représenté. Ainsi si cette plateforme est exploitée pour modéliser des molécules plus ou moins complexes, le volume représenté change, mais il correspond toujours au maximum à un cube de 1 000 000³ UV. Le principal atout de M-Cell est de gérer les aspects structurels des entités modélisées.

Ces plateformes dédiées ont été élaborées avec pour objectif de permettre de jouer le modèle pour lequel elles servent de support. Elles sont ainsi indissociables du modèle correspondant. Elles offrent peu de possibilités d'évolution et peu de souplesse, dans le sens où il est difficile de modifier ou d'adapter le modèle qu'elles vont jouer. Par exemple, pour modifier le modèle joué par INDISIM, les auteurs ont développé d'autres versions autour du socle de base que constitue INDISIM, INDISIM-SOM ou INDISIM-YEAST, et cette dernière a ensuite servi de base à l'élaboration d'INDISIM-Saccha.

Les objectifs de modélisation sont peu nombreux et ils sont définis en même temps que le modèle. Par exemple, SimCell ne permet pas la simulation de modèles qui décriraient l'adaptation d'une population bactérienne, sauf à modifier en profondeur le fonctionnement des entités qui sont modélisées et les règles d'interaction qui les concernent.

¹¹ <https://www.cellml.org/>

Ces plateformes se prêtent difficilement à la simulation de modèles multi-échelles. Elles sont prévues pour traiter des processus bien précis qui se déroulent généralement à une échelle donnée, par conséquent, la communication d'information entre les échelles, l'agrégation des informations d'un niveau particulier pour générer le comportement de niveau supérieur, etc... ne correspondent pas à des fonctions de ces plateformes.

III.4.3. PLATEFORMES GENERIQUES

À la différence des plateformes dédiées conçues pour un modèle spécifique, les plateformes génériques correspondent à des « framework » et permettent de faire jouer différents modèles tant qu'ils sont fournis selon le formalisme correspondant. Cependant si les modèles peuvent être différents, avec des objectifs différents, ces plateformes sont généralement conçues pour traiter un phénomène particulier. Ce dernier fait souvent appel à des notions multi-échelle et ces plateformes, dans leur conception, intègrent souvent la possibilité de faire communiquer les processus de différents niveaux, ou de faire jouer des modèles de niveau différents en parallèle.

Notons que si ces plateformes peuvent faire jouer plusieurs modèles en même temps, il ne s'agit pas d'une obligation, et qu'elles peuvent être utilisées pour simuler un seul modèle.

Ces plateformes offrent une certaine souplesse avec la possibilité de changer les modèles. La difficulté réside alors dans le besoin de maîtriser le formalisme pour exploiter le potentiel de la plateforme. Leur architecture est souvent plus complexe que celle des plateformes dédiées.

A. E-CELL

E-Cell est un projet commencé en 1996 au Laboratory for Bioinformatics de l'Université Shonan-Fujisawa de Keio, au Japon, sous le nom ECL pour Electronic Cell Laboratory. Il en est actuellement à sa version 3. E-Cell peut être trouvé en ligne ¹² (E-Cell, [117]).

Cette plateforme est à l'origine développée pour modéliser les processus biochimiques à l'œuvre dans une cellule. Ce niveau de description reste son objet d'étude préférentiel et lorsque les processus modélisés se déroulent à échelle cellulaire, le nombre de cellules limité.

L'objectif affiché par ses auteurs est de permettre la modélisation aussi précise que possible de l'ensemble des phénomènes à l'œuvre lorsqu'on observe une cellule, ainsi, ils ambitionnent de pouvoir décrire les aspects structurels des cellules et des molécules qu'elles contiennent, et la dynamique qui les régit en prenant en compte la physique et la chimie qui interviennent dans les réactions. La régulation de tous ces processus chimiques se fait par l'interaction avec l'environnement et se traduit par le contrôle de l'expression génétique des cellules et la dynamique des échanges entre les cellules et leur environnement. La capacité des cellules à capter des stimuli externes est aussi un élément important dans le développement des modèles simulés avec E-Cell.

¹² <http://www.e-cell.org/>

E-Cell permet de faire jouer des modèles écrits dans les langages SBML ou EML (E-Cell Model). E-Cell a par exemple permis de modéliser le métabolisme mitochondrial (Yugi et Tomita, 2004 [118]) et le métabolisme des globules rouges (Yachie-Konishita *et al.*, 2010 [119]) ou de rendre compte du rythme circadien chez la drosophile sous l'influence de la luminosité, et capable de rendre compte de l'effet de mutations sur les gènes impliqués (Smolen *et al.*, 2004 [120]).

B. COMPUCELL3D

CompuCell3D fait son apparition en 2003 avant d'être présenté officiellement en 2004 (Izaguirre *et al.*, 2004 [121]), mais il dérive des travaux de François Graner et James Glazier (CPM et Q-Potts) commencés en 1992 (Graner et Glazier, 1992 [42] ; Glazier et Graner 1993 [41]). Cette plateforme en est aujourd'hui à sa version 3.7.2 et est disponible en ligne ¹³ (CompuCell3D, [122]).

CompuCell3D est présenté par ses auteurs comme un support multi-modèle, il repose sur l'utilisation en parallèle de trois composantes, un modèle de Potts, un module gérant la diffusion et un module gérant des modèles mathématiques continus.

CompuCell3D est destiné à l'étude de la morphogénèse et s'appuie pour ça sur le savoir-faire des auteurs pour modéliser les réarrangements cellulaires dirigés par les forces qui s'exercent à leur surface et complété avec la prise en compte de la génétique des cellules et de la présence de substances chimiques comme des nutriments ou des signaux moléculaires dans l'environnement.

Cette plateforme nécessite des modèles écrits selon le langage CCML ou CC3DML, dérivés tous deux du standard XML. Le langage CC3DML est une extension du langage CCML qui permet la description des phénomènes dans 3 dimensions alors que CCML est limité à 2 dimensions.

CompuCell3D a fourni des résultats convaincants en morphogénèse en reproduisant des motifs caractéristiques de la croissance microbienne (Poplawski *et al.*, 2008 [123]) et en embryogénèse en permettant la simulation du processus de gastrulation chez l'embryon de poulet (Vasiev *et al.*, 2009 [124]) et la segmentation de l'embryon des vertébrés (Hester *et al.*, 2010 [125]; Dias *et al.*, 2014 [126]). Plusieurs utilisations probantes de cette plateforme ont été référencées par Maciej Swat (Swat *et al.*, 2012 [127]).

C. MORPHEX

Morphex est un projet de modélisation et de simulation de la morphogénèse et des réseaux de régulation génétique chez les plantes et animaux. Il a vu le jour en 2007 et s'intéresse tout particulièrement au développement des organes reproducteurs chez une espèce végétale, *Arabidopsis thaliana* et chez deux espèces d'éponges, *Suberites domuncula* and *Lubomirskia baikalensis*.

Le site internet de ce projet est disponible à l'adresse du site du projet ¹⁴ (Morphex, [128]).

¹³ <http://www.compuCell3d.org/>

¹⁴ <http://morphex.org/tiki-index.php?page=Home>

Le postulat de départ est que la croissance et le développement de l'organisme est présidé par le fonctionnement des cellules, lui-même contrôlé par l'expression du génome et l'influence que l'environnement exerce. Pour que le développement soit correct, il est important que le fonctionnement des cellules suive une chronologie précise, sinon des malformations apparaissent. L'objectif est ainsi d'identifier les aspects spatiaux et temporels qui gouvernent la mise en place des structures reproductrices des organismes étudiés, donc de décrire à l'échelle cellulaire le fonctionnement et les interactions cellulaires en fonction des signaux qu'elles captent et d'identifier l'ordre dans lequel tout ceci doit s'enchaîner pour que le développement soit correct. Les signaux que captent les cellules peuvent être émis par d'autres cellules ou être liés à l'environnement.

L'expérimentation *in silico* autorisant plus de liberté dans la manipulation des différentes variables et paramètres des expériences que ne le permettent les tests *in vitro* sans se heurter aux contraintes éthiques auxquelles peuvent se heurter les expériences *in vivo*. Ce projet s'appuie sur une méthodologie que les auteurs appellent COSMO pour COMplex System MOdeling, qui repose sur une plateforme du même nom et un langage dérivé du XML, le csmML (complex systems modeling Markup Language).

III.5. CONCLUSION

Cette description des modèles présents dans la littérature montre la difficulté à fédérer les efforts autour d'un même objectif de modélisation. Chaque modèle est élaboré par une équipe pour traiter une problématique. Chacun apporte sa vision des phénomènes importants à prendre en compte pour répondre à une question.

Nicholas Britton, dans son livre *Essential Mathematical Biology* de 2003 (Britton, 2003 [129]), a bien résumé le problème face auquel nous sommes confrontés, « tous les modèles ne seront pas accessibles à l'analyse mathématique, mais tous peuvent être simulés par ordinateur. Si la solution analytique existe, elle fournira une description complète [...] de la dynamique du système. Plusieurs modèles de simulation ne peuvent être décrits dans un cadre mathématique cohérent de sorte qu'ils soient accessibles à l'analyse mathématique. Ces modèles doivent être étudiés par l'analyse statistique d'un grand nombre de simulations. Le choix de l'approche du modèle dépend des caractéristiques du système dynamique lui-même et des aspects du système qui sont en accord avec les termes du modèle. Par ailleurs, les approches pluridisciplinaires sont essentielles car ceux qui sont experts dans un domaine particulier doivent travailler avec ceux qui sont experts des approches de modélisation mathématique ».

Il manque la dynamique pluridisciplinaire et la synergie d'équipes aux compétences complémentaires. Beaucoup de modèles qui ne traitent que de la croissance cellulaire se contentent de gérer la migration, la mort et la différenciation des cellules. Ce qui suffit pour observer la multiplication des cellules, mais ne permet pas de rendre compte des processus subcellulaires. A l'inverse, les modèles qui s'intéressent à cette échelle ne parviennent pas à rendre compte de l'aspect vivant d'une cellule, ils se contentent de la décrire comme un contenant dont la dynamique importe peu. Suivre une population de cellules sur une brève période de temps répond à la modélisation de phénomènes adaptatifs, et les modèles qui

assurent le suivi de la croissance cellulaire permettent de prendre en compte cette capacité des cellules à échanger de l'information avec les cellules qui les entourent ou avec leur environnement et à fournir une réponse adaptée. Cependant, la durée simulée par ces plateformes ne permet pas de fournir un modèle d'évolution, d'une part parce que le patrimoine génétique n'apparaît pas toujours explicitement, et qu'il n'est donc pas possible de le soumettre à un processus de sélection, et d'autre part parce qu'en regard de l'évolution, la vie d'une cellule représente trop peu. L'évolution ne peut apparaître qu'en mettant l'adaptation en perspective sur de nombreuses générations soumises à des contraintes fortes qui donnent prise à la sélection naturelle.

Or pour répondre à notre objectif radiobiologique, c'est-à-dire de suivre le comportement de cellules ou d'organismes à différentes échelles et de pouvoir les soumettre à un stress pour suivre leur réponse, stress suffisamment important pour induire un processus de sélection qui fera évoluer la population suivie, il nous est nécessaire de disposer d'un modèle capable de prendre en charge ces différentes échelles temporelles et spatiales.

Si l'objectif à long terme est de disposer d'un tel modèle suffisamment générique pour modéliser des cellules, des organes ou des organismes complexes complets, le premier jalon à atteindre est de pouvoir fournir un tel modèle capable de rendre compte de tous ces processus à l'échelle des organismes parmi les plus simples, les bactéries. Les bactéries offrent un double avantage. D'une part leur structure simple rend l'élaboration du modèle moins complexe, et d'autre part, du fait que leur manipulation est simple et rapide, il est facile de produire des résultats expérimentaux. Ces derniers peuvent alors simplement être confrontés à ceux fournis par les simulations qui peuvent être réalisées à l'aide de ce modèle.

Le modèle dont nous aurons besoin semble ne pas exister et il nous faut l'élaborer. Pour assurer sa simulation, on l'a vu, les plateformes dédiées ne se prêtent pas à l'exercice de faire jouer un modèle différent de celui pour lequel elles ont été conçues. On pourrait alors se tourner vers les plateformes génériques, mais on s'aperçoit rapidement que dans leur conception, même si elles présentent une capacité multi-échelle, cet aspect reste limité à deux échelles qui en plus sont gérées par des modèles différents et indépendants. Faire transiter de l'information entre les échelles apparaît de fait complexe à réaliser.

Au final, après cet état de l'art, il nous apparaît nécessaire d'élaborer un nouveau modèle pour le fonctionnement des cellules, sans perdre de vue la perspective radiobiologique. Ainsi notre modèle sera un modèle individu-centré, et il devra gérer les processus biochimiques à l'échelle subcellulaire, le fonctionnement et la vie des cellules à l'échelle des cellules, en conservant leur patrimoine génétique pour assurer le suivi des populations qui peuvent apparaître sous les pressions de sélection de l'environnement (Tableau III.1).

En parallèle, il nous faut développer la plateforme qui permettra de simuler ce modèle. Il nous apparaît en effet plus simple de construire la plateforme au fur et à mesure que le besoins se feront jour dans le modèle, plutôt que d'envisager *ab initio* tous les aspects fonctionnels requis, ou de reconstruire les modules répondant à nos besoins dans les plateformes génériques existantes. Nous envisageons toutefois de garder en perspective dans la conception de notre plateforme un aspect générique, pour permettre l'utilisation de

plusieurs modèles si le besoin s'en fait sentir, ou pour permettre sa réutilisation dans un autre cadre.

Nous avons vu que les automates cellulaires, au sens large, constituent un modèle de choix pour modéliser les systèmes complexes, leur analogie de fonctionnement avec le vivant à l'échelle cellulaire rend de plus la traduction de processus biologiques en règles de transition relativement intuitive, aussi faisons-nous ce choix pour le développement de notre modèle et de notre plateforme.

	Modèle/plateforme dédiée										Plateformes génériques		Problématique	
	Avida			BacSim			SimCell		CompuCell3D		Règles	Observables	Règles	Observables
	Règles	Observables		Règles	Observables		Règles	Observables	Règles	Observables				
Echelle spatiale	Organisme								X	X				
	Organe								X	X				
	Tissu/Colonie	X			X				X	X			X	X
	Cellule	X		X	X				X	X			X	X
	Molécule						X	X	X	X			X	X
	Atome						X	X	X	X			X	X
Echelle temporelle	Génération	X	X		X				X	X			X	X
	Vie de l'individu			X	X				X	X			X	X
	Réaction chimique						X	X	X	X			X	X
Changement	Réversible				X					X			X	
	Héréditaire	X			X								X	
Echanges entre échelles		Oui	Cellule --> Population	Oui	Cellule <--> Population	Oui	Atome <--> Molécule	Non, à l'exception de cellule --> tissu			Oui, bidirectionnelle entre toutes les échelles			

Tableau III.1 : Récapitulatif des propriétés de quelques unes des plateformes étudiées par rapport à nos besoins.

CHAPITRE IV. PROPOSITIONS : UNE PLATEFORME, UN MODELE DE CELLULE BIOLOGIQUE

IV.1. Plateforme	113
IV.1.1. Présentation	113
IV.1.1.1. Cahier des charges et hypothèses	113
IV.1.1.2. Architecture	114
IV.1.2. Applications : modèles et validation de la plateforme	116
IV.1.2.1. Adaptations de la plateforme liées à l'automate cellulaire	116
IV.1.2.2. Jeu de la vie	118
IV.1.2.2.1. Spécifications	118
IV.1.2.2.2. Analyse	119
IV.1.2.2.3. Implémentation	121
IV.1.2.2.4. Mise en œuvre	122
IV.1.2.3. Wireworld	122
IV.1.2.3.1. Spécifications	122
IV.1.2.3.2. Analyse	123
IV.1.2.3.3. Implémentation	125
IV.1.2.3.4. Mise en œuvre	126
IV.1.2.4. Daisyworld	127
IV.1.2.4.1. Spécification	127
IV.1.2.4.2. Simplification de Daisyworld pour l'étude de l'homéostasie en système isolé	128
IV.1.2.4.3. Analyse	130
A. Version simplifiée de Daisyworld pour l'étude de l'homéostasie	131
B. Version classique de Daisyworld	132
IV.1.2.4.4. Implémentation de Daisyworld	135
IV.1.2.4.5. Mise en œuvre de Daisyworld	135
IV.1.2.5. Lien avec d'autres travaux de l'équipe	137
IV.1.3. Remarques sur les temps de calculs	138
IV.1.4. Conclusion	140
IV.2. Modèle de cellule	141
IV.2.1. Hypothèses de modélisation	142
IV.2.1.1. Fonctionnement de la cellule	142
IV.2.1.2. Cycle de vie de la cellule	144
IV.2.2. Paramètres du modèle	146
IV.2.3. Devenir des cellules filles	150
IV.2.4. Superindividu	150
IV.3. Définition des valeurs des paramètres	152
IV.3.1. Consommation par bactérie	153
IV.3.2. Environnement modélisé	153
IV.3.3. Densité de population bactérienne donnant une DO de 1	154

IV.4. Fonctionnement et implémentation	155
IV.4.1. Paramètres de l'automate cellulaire.....	155
IV.4.2. Gestion et suivi de la simulation.....	155
IV.4.3. Initialisation d'une cellule.....	157
IV.4.3.1. Initialisation d'une cellule de la génération 0	157
IV.4.3.2. Initialisation à la naissance d'une cellule	157
IV.4.4. Fonctions cellulaires.....	158
IV.4.4.1. Fonction régissant le déroulement du cycle cellulaire	158
IV.4.4.2. Fonction régissant les échanges entre la cellule et son environnement au cours du processus de division.....	159
IV.5. Conclusion.....	161

Lors de la description des systèmes biologiques, nous avons identifié les caractéristiques que nous avons besoin de prendre en compte pour répondre à notre problématique. Nous avons ainsi dressé un cahier des charges du modèle dont nous avons besoin. En le rappelant et en le comparant à l'existant, nous arriverons à la proposition d'un modèle. Mais auparavant, nous avons besoin d'un environnement logiciel sur lequel exploiter ce modèle.

IV.1. PLATEFORME

Pour faire jouer notre modèle une plateforme a été développée en interne à l'équipe, pour avoir une plus grande liberté d'implémentation par rapport aux plateformes décrites précédemment. Nous avons ainsi pu garder la main sur la gestion des fonctionnalités, la plateforme ayant pu être enrichie à mesure de l'élaboration du modèle.

IV.1.1. PRESENTATION

L'objectif est de proposer un environnement logiciel simple à manipuler qui permettra de mettre en œuvre des modèles de population cellulaire.

Valider le comportement de la plateforme et vérifier qu'elle ne s'accompagne pas de biais dans sa conception impose de reproduire des modèles existants et comparer les résultats que nous obtenons à ceux qui sont connus pour ces modèles. Cette plateforme doit donc être assez souple pour permettre aussi facilement la mise en œuvre de nouveaux modèles que de modèles de la littérature.

IV.1.1.1. CAHIER DES CHARGES ET HYPOTHESES

Pour répondre au besoin de modularité et faciliter la mise en œuvre de modèles, qu'ils soient anciens ou nouveaux, l'architecture de la plateforme devra être générale. Sa structure devra être la plus simple possible pour que la mise en œuvre de modèles se fasse sans un gros travail d'écriture, même si elle nécessite quelques connaissances en informatique.

Le fonctionnement de la plateforme ne doit pas induire de comportement particulier ni d'exécution particulière, tout doit dépendre du modèle. L'architecture devra ainsi éviter des biais, dans l'exécution ou dans l'écriture du modèle. Les limites rencontrées doivent être le fait du modèle et non de l'environnement logiciel, c'est-à-dire que le fonctionnement du modèle ne doit pas être contraint par la plateforme, que ce soit en termes de temps, d'espace, ou autre. Le temps et l'espace seront donc des propriétés des modèles et non de la plateforme.

Nous avons vu lors de la description des systèmes biologiques qu'ils sont par nature multi-échelles. Modéliser une population cellulaire doit pouvoir se faire à plusieurs niveaux. Nous nous fixons donc comme objectif de pouvoir utiliser des modèles multi-échelles, ou à tout le moins de faire du multi-échelle, en faisant cohabiter des modèles différents pour les différentes échelles.

Notre attention se portant principalement sur les populations cellulaires, ce sera donc l'objet des modèles individu-centrés que nous étudierons. Nous choisissons de modéliser les individus, et par conséquent les cellules, par deux objets qui décriront d'une part les processus, et d'autre part les caractéristiques. Il n'y aura ainsi pas de classe spécifique pour modéliser les cellules biologiques, mais deux classes, une structure qui est manipulée par un modèle.

Les différents modèles qui auront été implémentés seront stockés et disponibles dans une bibliothèque de modèles qui permettra leur réutilisation ultérieure, soit pour combiner leurs approches, soit pour les présenter facilement à titre d'illustration par exemple.

IV.1.1.2. ARCHITECTURE

Les différents niveaux de modélisation seront associés à des classes. Le niveau microscopique sera adressé par la classe *CellModel*, le niveau macroscopique sera adressé par la classe *Population*. Produire un modèle individu-centré se fera en dérivant la classe *CellModel*, alors que produire un modèle à population se fera en dérivant la classe *Population*. Le niveau de modélisation mésoscopique se fera en définissant une population associée à un modèle qui représentera un groupe d'individus et non plus un seul individu.

Cette architecture est présentée à la Figure IV.1.

L'élément central de l'architecture sera la classe *Population*. C'est à son niveau que se fera la gestion des individus et leurs interactions avec leur environnement. Ces interactions seront définies au niveau du *CellModel*, mais le passage des informations entre les instances de *CellModel* et l'environnement sera effectué par la *Population*. Elle permettra également de propager entre les individus les informations de mise à jour liées aux relations interindividuelles.

De fait, c'est la définition de la population qui indiquera le type de modélisation (discrète, continue, contrôlée par horloge...). La *Population* manipule un ensemble de *CellModel*. Elle peut être homogène et présenter plusieurs instances d'un même *CellModel*, ou hétérogène avec plusieurs instances de différents *CellModel*.

Un *CellModel* correspond à un type de cellule et ne manipule donc qu'une seule structure. Chaque individu est une instance d'un *CellModel* qui manipule une instance d'une *CellStructure*.

Les attributs de base d'une *CellStructure*, présents dans la classe abstraite *CellStructure*, correspondent à ses coordonnées cartésiennes, *_IdX* et *_IdY*). Comme il n'y a pas d'hypothèse sur l'espace dans la plateforme, ces attributs sont valables que l'on considère un espace continu ou discret, à une ou deux dimensions.

La démarche pour implémenter un nouveau modèle consiste à créer les classes dérivées des classes abstraites *CellStructure*, qui porte les attributs représentant les caractéristiques biologiques modélisées, et *CellModel* qui présente les méthodes représentant les processus biologiques d'intérêt (le cycle cellulaire en particulier pour un modèle individu-centré de cellule). S'il est besoin d'un environnement particulier, il sera décrit dans la classe

Environment et présentera des attributs qui représenteront toutes les caractéristiques d'intérêt pour le modèle, comme par exemple une information de température, la quantité d'un nutriment disponible, etc... Les attributs de cette classe peuvent être consultés ou modifiés par les *CellModel* en passant par la *Population*.

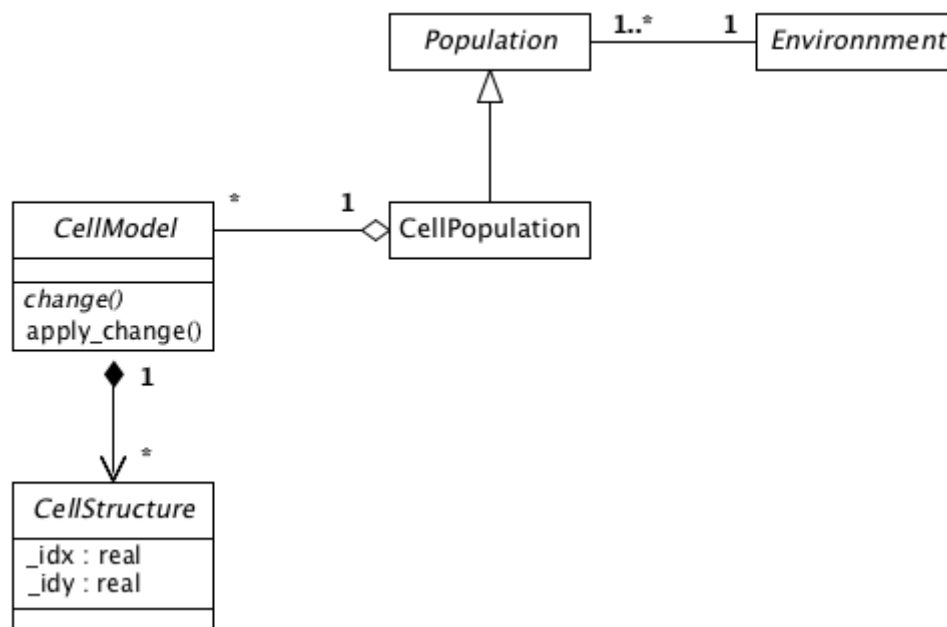


Figure IV.1 : Diagramme de classe de l'architecture de la plateforme.

La classe *CellModel* présente une méthode abstraite, « *Change ()* » et une méthode concrète « *ApplyChange ()* ». Cette dernière n'a d'intérêt que pour les modèles individu-centrés. Elle applique à l'individu le nouvel état qui est calculé par la méthode « *Change ()* », elle est donc indépendante du modèle, à la différence de la méthode « *Change ()* » qui doit être rendue concrète dans la classe concrète *CellModel*. C'est donc à la définition de cette dernière en dérivant la classe *CellModel* que sera défini le contenu de la méthode « *Change ()* ».

En résumé, cette architecture permet la manipulation concrète au niveau d'une classe abstraite, la *CellStructure* dérivée de *CellStructure* est manipulée par la méthode « *Change ()* » du *CellModel* dérivé de *CellModel*.

On donnera le choix à l'utilisateur de spécifier en début d'exécution les conditions initiales ou de laisser calculer une configuration aléatoire, par exemple, pour jouer un modèle de population bactérienne, soit l'utilisateur fournit une population donnée, soit la plateforme en génère une aléatoirement.

L'environnement pris en compte dans la classe *Environment* est actuellement simple, mais il sera à terme capable de proposer des listes de paramètres chimiques, par exemple des listes de nutriments et leur quantité respective, une liste de toxiques, etc... et des paramètres physiques comme la température, la pression, etc... Chacun de ces éléments doit pouvoir être rendu disponible et consulté par les individus.

IV.1.2. APPLICATIONS : MODELES ET VALIDATION DE LA PLATEFORME

À l'issue de notre état de l'art, nous avons choisi de nous orienter vers l'utilisation des automates cellulaires comme support de modèles individu-centrés. Aussi, nous nous sommes intéressés à implémenter dans la plateforme des automates cellulaires bien connus dans la littérature, le Jeu de la Vie et Wireworld.

Nous avons également choisi d'expérimenter l'implémentation d'un modèle d'homéostasie*, Daisyworld, sous la forme d'automates cellulaires en considérant d'abord une version simplifiée avec un système isolé, puis la version classique avec un système fermé.

Nous orienter vers les automates cellulaires pour valider le comportement de la plateforme et mettre en œuvre notre modèle de cellules biologiques nous a imposé certaines adaptations ou certaines précisions sur elle.

IV.1.2.1. ADAPTATIONS DE LA PLATEFORME LIEES A L'AUTOMATE CELLULAIRE

La définition d'un automate cellulaire donne une fonction de transition qui régit le devenir de l'état d'une cellule dans le temps à travers l'interaction avec les cellules considérées dans son voisinage. L'utilisation de la plateforme pour mettre en œuvre ces automates cellulaires passe donc par la définition de leur fonction de transition au sein de la méthode « Change() » de la classe *CellModel* développée pour chacun d'eux.

Avec la fonction de transition, il faut définir l'ensemble des états que peut prendre chaque cellule, ces états seront représentés par l'attribut `_state` de la classe *CellStructure*.

La fonction de transition d'un automate cellulaire permet de calculer l'état d'une cellule à un temps $t+1$ à partir de son état à un temps t et de l'état de ses cellules voisines. Il nous faut donc développer une classe qui contient la méthode de calcul du voisinage.

C'est également à son niveau que se décide la topologie de l'espace (1D, 2D, ...), si l'espace est continu ou discret, et dans ce dernier cas, la forme du maillage. Nous avons implémenté le calcul des voisinages de Von Neumann et de Moore, avec pour ce dernier plusieurs rayons possibles (Figure IV.2). Nous considérons qu'un voisinage de Moore de rayon infini correspond à un voisinage global, c'est-à-dire à l'utilisation de la matrice entière comme système de voisinage pour chaque cellule.

Cette classe est sollicitée lors de l'initialisation de l'automate cellulaire et à chaque fois que les voisins ont besoin d'être recalculés, par exemple si l'on considère que les individus se déplacent.

Nous avons implémenté cinq processus de mise à jour du modèle. Ils constituent des méthodes de la classe *Population*.

* L'homéostasie est la capacité d'un système à maintenir son équilibre de fonctionnement en dépit de contraintes extérieures

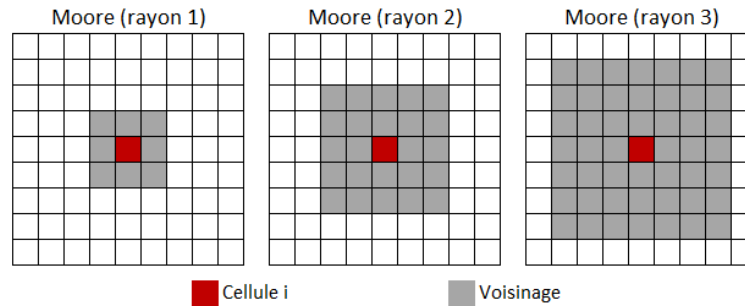


Figure IV.2 : Trois exemples de voisinage de Moore, respectivement de rayon 1, 2 et 3.

« UpdateAll » décrit une mise à jour synchrone du système. Tous les éléments sont au même pas de temps. Cette méthode appelle la méthode « Change » de chaque CellModel à chaque pas de temps de la simulation, et appelle la méthode « ApplyChange () » de CellModel à la fin de chaque pas de temps. « Change () » calcule l'état à $t+1$ en à partir de l'état à t et en fonction des voisins, et tous sont encore à l'état t . L'état à $t+1$ est appliqué à tous les individus en même temps. Avec cette méthode, les cellules sont synchrones entre les pas de temps et dans le pas de temps. Ce processus a été implémenté de deux façons, la première, « UpdateAll () », réalisant la mise à jour des cellules dans leur ordre lexicographique (*i.e.* dans l'ordre de leurs coordonnées dans l'espace), la seconde, « UpdateAllRand () », procédant à la mise à jour des cellules dans un ordre aléatoire.

« UpdateAllAsynchrone » décrit une mise à jour asynchrone-synchrone du système. Au cours d'un pas de temps de la simulation, l'état à $t+1$ d'un CellModel est calculé par appel de sa méthode « Change () ». Ce nouvel état est aussitôt appliqué par appel de la méthode « ApplyChange () ». Ainsi, le calcul de l'état d'une cellule prend en compte son état à l'instant t , et l'état de ses voisines. Mais comme certaines ont été mises à jour, certaines sont à l'état t , d'autres sont à l'état $t+1$. Le pas de temps est terminé lorsque toutes les cellules ont été mises à jour et que toutes sont à l'état $t+1$. En résumé, les cellules sont synchrones entre les pas de temps, mais asynchrones au cours d'un pas de temps, la resynchronisation ayant lieu à la fin de chaque pas de temps. Ce processus a également été implémenté sous deux formes, « UpdateAllAsynchrone () » réalisant la mise à jour des cellules dans leur ordre lexicographique, et « UpdateAllAsynchroneRand () » procédant à la mise à jour des cellules dans un ordre aléatoire.

« UpdateAllAsynchroneOne () » reprend le principe de la méthode « UpdateAllAsynchroneRand () » mais avec la disparition de la notion de pas de temps de la simulation. Ou plus exactement, le calcul de l'état à $t+1$ et l'application de ce nouvel état à une cellule compte pour un pas de temps. De fait, avec un ordre de mise à jour aléatoire des cellules, certaines seront mises à jour pour la première fois alors que certaines de leurs voisines auront déjà subies plusieurs mise à jour, les cellules sont toutes désynchronisées et le processus est totalement asynchrone. Ce processus utilise forcément un ordre de mise à jour des cellules aléatoire et n'a donc été implémenté que sous cette seule forme.

Pour les méthodes qui permettent la mise à jour des cellules dans un ordre aléatoire, nous avons fait l'essai de mettre calculer l'ordre de mise à jour une fois au début de l'exécution, ou à chaque itération. Les résultats de ces deux façons de procéder sont comparés (voir le paragraphe V.2, page 167).

Le principe de ces processus de mise à jour est résumé et comparé dans le Tableau IV.1

UpdateAll ()	UpdateAllAsynchrone ()	UpdateAllAsynchroneOne ()
Pour chaque cellule <i>i</i> faire Change () Fin pour Pour chaque cellule <i>i</i> faire ApplyChange () Fin pour	Pour chaque cellule <i>i</i> faire Change () ApplyChange () Fin pour	Tant que la simulation Choisir une cellule Change () ApplyChange () Fin tant que

Tableau IV.1 : Comparaison des algorithmes de mise à jour implémentés.

IV.1.2.2. JEU DE LA VIE

IV.1.2.2.1. SPECIFICATIONS

Le Jeu de la Vie a été proposé par John Conway en 1970 sous le nom Game of Life. Il s'agit de l'un des automates cellulaires les plus documentés depuis sa popularisation par Martin Gardner qui le présente dans son article «The fantastic combinations of John Conway's new solitaire game "life"» (Gardner, 1970 [130]).

Il s'agit d'un automate cellulaire déterministe et homogène, synchrone, basé sur un maillage régulier exploitant un voisinage de Moore.

Les cellules peuvent présenter deux états :

- active (ou vivante) ;
- inerte (ou morte).

Le comportement de cet automate cellulaire est dicté par une fonction de transition comprenant trois règles :

- Une cellule morte devient vivante si deux ou trois cellules sont vivantes dans le voisinage, elle reste morte sinon ;
- Une cellule vivante meurt d'isolement si elle a au plus une cellule vivante dans le voisinage ;
- Une cellule vivante meurt par surpopulation si elle a dans son voisinage au moins quatre cellules vivantes.

Cette fonction peut aussi s'écrire de manière suivante :

- Une cellule devient ou reste active si elle a dans son voisinage deux ou trois cellules actives, dans tous les autres cas, elle devient ou reste inactive.

Cette fonction simple permet de générer des comportements relativement complexes. Des conditions initiales aléatoires peuvent faire apparaître plusieurs structures remarquables regroupées dans trois classes : oscillateurs, stables et planeurs (ou gliders).

Les oscillateurs sont des structures qui possèdent plusieurs formes et passent de l'une à l'autre régulièrement avec une périodicité caractéristique. L'oscillateur le plus simple, la structure en haut à gauche dans la Figure IV.3, est constitué d'un alignement de trois cellules actives et bascule entre chaque forme à chaque pas de temps.

Les formes stables gardent la même structure dans le temps. La plus petite forme stable, représentée en bas à droite dans la Figure IV.3, est constituée d'un carré de quatre cellules actives.

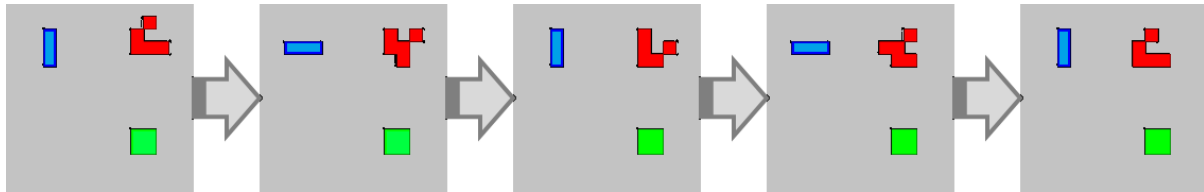


Figure IV.3 : Évolution sur 5 pas de temps de figures simples observées dans le Jeu de la Vie. L'oscillateur, en haut à gauche, alterne ses deux formes à chaque pas de temps ; la forme stable, en bas à droite, reste identique d'un pas de temps sur l'autre ; le planeur, en haut à droite, se retrouve à l'identique après 4 pas de temps, mais décalé dans l'espace.

Les planeurs sont des structures qui évoluent dans le temps et se retrouvent à l'identique mais déplacés dans la matrice au bout de quelques pas de temps. Le plus simple, en haut à droite dans la Figure IV.3, se déplace en diagonale de trois cases tous les quatre pas de temps. Il existe beaucoup de planeurs différents caractérisés par leur orientation et la direction de leur déplacement, leur périodicité et leur vitesse.

Des conditions initiales aléatoires vont produire des comportements à l'allure chaotique dont surgissent rapidement des formes pouvant appartenir aux classes décrites précédemment.

Certaines conditions initiales ont été identifiées et caractérisées pour l'évolution qu'elles suivent et le motif qu'elles dessinent dans le temps ou à un instant donné, par exemple, le clown, qui apparaît à la 111^{ème} itération de l'automate à partir d'une structure initiale en forme de « u » inversé constitué de sept cellules actives (Figure IV.5).

Il existe des recherches pour inventorier les planeurs, oscillateurs ou motifs stables qui existent, ainsi que les conditions initiales au comportement remarquable. Saisir les mots clés « game of life » ou « jeu de la vie » sur un moteur de recherche renvoie de nombreux liens vers des bibliothèques répertoriant les configurations initiales particulières connues ou les structures que présente le Jeu de la vie.*

L'objectif à travers l'implémentation de cet automate cellulaire est de vérifier qu'à l'aide de la plateforme, nous sommes capables d'obtenir les formes présentes dans le jeu de la vie (planeurs, formes stables et oscillateurs) et de reproduire des configurations remarquables identifiées dans la littérature telles que des canons à planeurs ou le clown.

IV.1.2.2.2. ANALYSE

Pour implémenter ce modèle au sein de la plateforme, nous avons défini trois données par cellule :

* Exemples de sites présentant les configurations remarquables et structures connues dans le jeu de la vie :

<http://www.math.com/students/wonders/life/life.html>

<http://conwaylife.com/>

<https://www.wolframscience.com/nksonline/page-964f-text?firstview=1>

- $_state_i \in \{0; 1\}$ pour représenter l'état actuel de la cellule i , inactive ou active respectivement ;
- $_futureState_i \in \{0; 1\}$ pour représenter l'état calculé de la cellule i , inactive ou active respectivement ;
- N_i = Ensemble des cellules comprises dans le voisinage de Moore de la cellule i .

Ainsi qu'une variable locale

- Somme $\in \mathbb{N}$, une somme de valeurs d'état.

L'algorithme de mise à jour de cet automate s'écrit alors (Algorithme IV.1) :

```

Début
|   Pour chaque cellule  $i$  faire
|   |   Somme  $\leftarrow 0$ 
|   /   /   ** On compte le nombre de cellules actives dans le voisinage de la cellule  $i$  **
|   |   Pour chaque cellule  $n$  de  $N_i$  faire
|   |   |   Somme  $\leftarrow$  Somme +  $\_state_n$ 
|   |   Fin pour
|   |   Si Somme = 2 ou Somme = 3 alors
|   |   /   /   /   ** La cellule  $i$  est active **
|   |   |    $\_futureState_i \leftarrow 1$ 
|   |   Sinon
|   |   /   /   /   ** La cellule  $i$  est inactive **
|   |   |    $\_futureState_i \leftarrow 0$ 
|   |   Fin si
|   Fin pour
|   Pour chaque cellule  $i$  faire
|   |    $\_state_i \leftarrow \_futureState_i$ 
|   Fin pour
Fin

```

Algorithme IV.1 : Algorithme de principe du Jeu de la Vie.

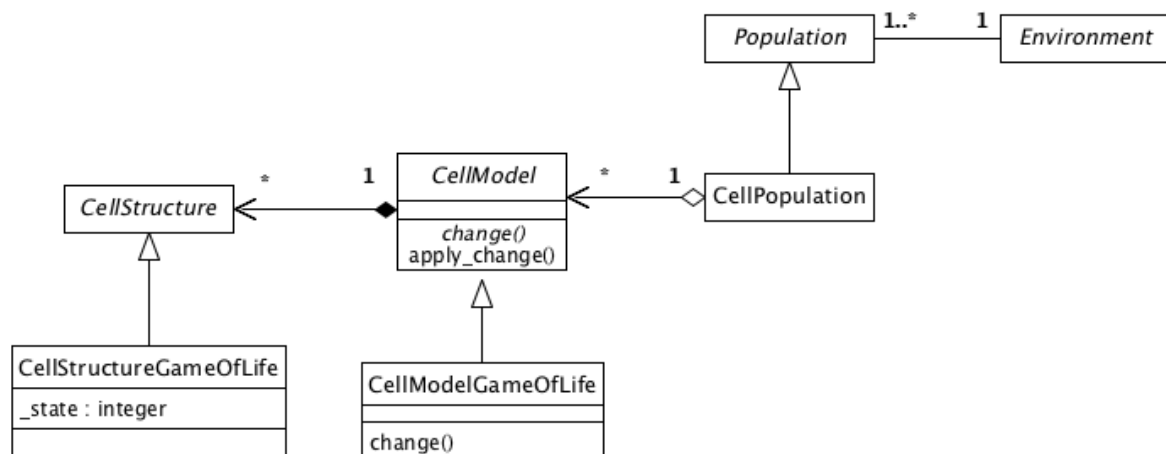


Figure IV.4 : Diagramme de classe de l'implémentation du Jeu de la Vie au sein de la plateforme.

Les données des cellules sont stockées dans les attributs de la classe CellStructureGameOfLife dérivée de CellStructure (Figure IV.4).

Le voisinage N_i n'apparaît pas dans le diagramme de classes, il est calculé au niveau de la structure de données qui stocke la population, les individus interrogent la population à laquelle ils appartiennent pour avoir la connaissance de ce N_i . Il en est de même pour la variable `_futureStatei` qui est stockée par la population pour assurer la mise à jour des cellules.

L'algorithme représentant la fonction de transition de cet automate cellulaire et les variables locales s'insèrent dans la classe `CellModelGameOfLife` dérivée de *CellModel* (Figure IV.4).

IV.1.2.2.3. IMPLEMENTATION

L'implémentation passe par la conception des classes correspondant au modèle. Elles sont au nombre de trois :

- la classe `fileReader` qui permet de spécifier la configuration initiale qui va être employée pour la simulation ;
- la classe `CellStructureGameOfLife` qui intègre les variables représentant les cellules du Jeu de la vie ;
- la classe `CellModelGameOfLife` qui contient la fonction de transition de l'automate.

Pour donner un exemple, l'algorithme de la fonction de transition est implémenté dans la méthode « `Change ()` » de la classe `CellModelGameOfLife` (Algorithme IV.2)

```
public void change() {
    int laSomme = 0;
    GameOfLifeStructure leNewState = new GameOfLifeStructure(this.get_status().get_idx(),
this.get_status().get_idy());
    ** On parcourt les voisins de la cellule courante **
    for (CellModel mdl : this._voisins) {
        ** On fait la somme des valeurs d'états des cellules voisines **
        if (((GameOfLifeStructure) mdl.get_status()).get_status()) {
            laSomme++;
        }
    }
    if ((laSomme >= 2) && (laSomme <= 3)) {
        ** Si la cellule courante a 2 ou 3 voisines actives, elle sera active à la prochaine itération **
        leNewState.set_status("1"); // 2 ou 3 voisines vivantes => vivante
    } else {
        ** Sinon la cellule courante sera inactive à la prochaine itération **
        leNewState.set_status("0"); // - de 2 ou + de 3 voisines vivantes => morte
    }
    this.set_New_Status(leNewState);
}
```

Algorithme IV.2 : Algorithme de l'implémentation du Jeu de la Vie.

L'application de la nouvelle valeur d'état à la valeur d'état de chaque cellule n'apparaît pas dans cette méthode, elle est réalisée par la méthode de mise à jour de l'automate « `UpdateAll ()` » de la classe *Population*.

IV.1.2.2.4. MISE EN ŒUVRE

L'implémentation de cet automate a produit les résultats conformes à ceux attendus. À partir de conditions initiales aléatoires, le comportement de cet automate est chaotique avant de se stabiliser rapidement et de présenter un environnement peuplé de formes simples (Figure IV.3). Des configurations initiales précises inspirées de la littérature et ont également été reproduites avec succès, nous illustrerons l'exemple avec la configuration du clown (Figure IV.5).

Le résultat d'une simulation de cet automate peut être visualisé en temps réel à l'écran à travers une interface graphique, ou en générant itération par itération un fichier image présentant la configuration de l'automate.

La bonne reproduction de cet automate cellulaire et de son comportement a permis à la plateforme de démontrer sa bonne gestion d'un automate cellulaire binaire avec des règles déterministes et homogènes prenant en compte un voisinage de Moore, dans un réseau régulier de cellules carrées.

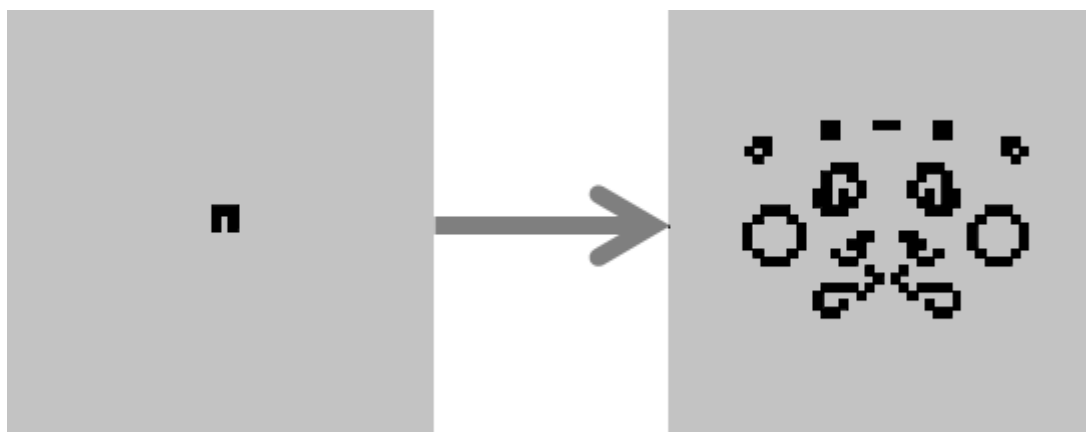


Figure IV.5 : Le clown apparaît à la 111^{ème} itération du Jeu de la Vie.

IV.1.2.3. WIREWORLD

IV.1.2.3.1. SPECIFICATIONS

Wireworld est un automate cellulaire qui a été décrit par Brian Silverman en 1987 et popularisé par A.K. Dewdney en 1990 (Dewdney, 1990 [131]) dont le but est de reproduire le comportement d'un électron dans un fil électrique. Il s'agit d'un automate cellulaire déterministe utilisant un réseau régulier de mailles carrées. Les règles de transition sont homogènes et synchrones. Elles reposent sur l'exploitation du voisinage de Moore.

Les règles fonctionnelles de cet automate décrivent trois états:

- Fil électrique ;
- Électron ;
- Trace ;

La fonction de transition qui régit le comportement de cet automate cellulaire comprend trois règles :

- un électron devient une trace ;
- une trace devient un fil électrique ;
- un fil devient un électron s'il y a dans son voisinage un ou deux électrons ;

Puisque l'on va chercher à tracer des circuits électriques et à créer des composants électroniques, on peut éventuellement rajouter un quatrième état « vide » qui traduira l'espace entre deux fils. La fonction de transition peut par conséquent être complétée par une quatrième règle liée à cet état : une maille vide restera toujours vide.

Cette fonction de transition permet de reproduire le comportement de courants électriques dans des fils électriques. En assemblant judicieusement les fils électriques il devient possible de reproduire des composants électroniques simples, tels des diodes, des portes logiques, des générateurs. De fait cet automate cellulaire a été exploité pour reproduire des structures électroniques complexes allant même jusqu'à simuler des ordinateurs capables de calculer les nombres premiers (quinapaslu [132]).

Nous n'avons pas poussé à ce point l'exploitation de cet automate cellulaire, mais nous avons réalisé des circuits simples incluant certains des composants de base (Figure IV.7), notre objectif étant de reproduire le comportement de l'automate cellulaire sur des cas simples pour vérifier la gestion d'automates à plus de deux états.

IV.1.2.3.2. ANALYSE

Pour implémenter ce modèle au sein de la plateforme, nous avons défini deux variables globales :

- $_state_i \in \{\text{Fil} ; \text{Électron} ; \text{Trace} ; \text{Vide}\}$ pour représenter l'état courant de la cellule i ;
- $_futureState_i \in \{\text{Fil} ; \text{Électron} ; \text{Trace} ; \text{Vide}\}$ pour représenter l'état de la cellule i à l'itération suivante ;
- N_i = Ensemble des cellules comprises dans le voisinage de Moore de la cellule i .

Ainsi qu'une variable locale

- $nbE \in \mathbb{N}$, un nombre d'électron.

L'algorithme de mise à jour de cet automate s'écrit alors (Algorithme IV.3) :

```

Début
|   Pour chaque cellule  $i$  faire
|   /   ** Si la cellule  $i$  est un électron, elle devient une trace **
|   |   Si  $\_state_i = \text{Électron}$  alors
|   |   |    $\_futureState_i \leftarrow \text{Trace}$ 
|   /   ** Sinon si la cellule  $i$  est une trace, elle devient un fil **
|   |   Sinon si  $\_state_i = \text{Trace}$  alors
|   |   |    $\_futureState_i \leftarrow \text{Fil}$ 
|   /   ** Si la cellule  $i$  est un fil **
|   |   Sinon si  $\_state_i = \text{Fil}$ 
|   |   |    $nbE \leftarrow 0$ 
|   |   |   Pour chaque cellule  $n$  de  $N$  faire
|   /   /   /   ** On compte le nombre d'électrons présents dans son voisinage **

```

```

|         |         |         |         Si _staten = Électron alors
|         |         |         |         |         nbE ← nbE + 1
|         |         |         |         Fin si
|         |         |         Fin pour
|         |         /         /         /         ** Si le nombre d'électron dans le voisinage de la cellule i est 1 ou 2 **
|         |         |         |         Si nbE = 1 ou nbE = 2 alors
|         |         |         |         |         ** La cellule i, devient un électron **
|         |         |         |         |         _futureStatei ← Électron
|         |         |         |         Sinon
|         |         |         |         |         ** La cellule i reste un fil **
|         |         |         |         |         _futureStatei ← Fil
|         |         |         |         Fin si
|         |         /         /         ** Sinon, une maille vide reste vide **
|         |         |         Sinon
|         |         |         |         _statei ← Vide
|         |         |         Fin si
|         Fin pour
|         Pour chaque cellule i faire
|         |         _statei ← _futureStatei
|         Fin pour
Fin

```

Algorithme IV.3 : Algorithme de principe de Wireworld.

Les données des cellules sont utilisées dans la classe CellStructureWireworld dérivée de CellStructure (Figure IV.6).

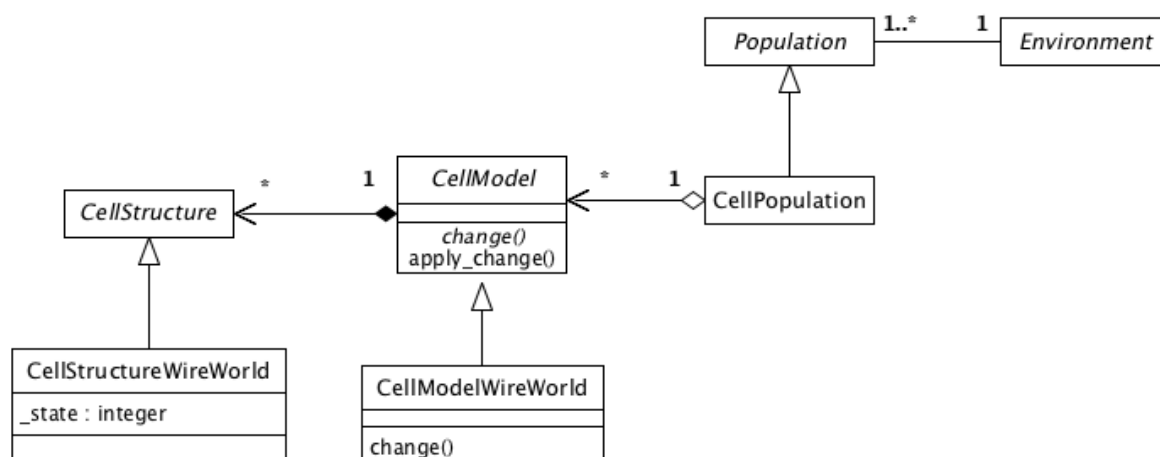


Figure IV.6 : Diagramme de classe de l'implémentation de Wireworld au sein de la plateforme.

Le voisinage N_i n'apparaît pas dans le diagramme de classes, il est calculé au niveau de la structure de données qui stocke la population, les individus interrogent la population à laquelle ils appartiennent pour avoir la connaissance de ce N_i . De même la valeur $_futureState_i$ est manipulée au niveau de la classe population pour permettre la mise à jour de l'état de chaque cellule.

L'algorithme représentant la fonction de transition de cet automate cellulaire et les variables locales s'insèrent dans la classe CellModelWireworld dérivée de *CellModel* (Figure IV.6).

IV.1.2.3.3. IMPLEMENTATION

L'implémentation passe par la conception des classes correspondant au modèle. Elles sont au nombre de trois :

- la classe FileReader qui permet de spécifier la configuration qui va être employée pour la simulation, soit le « dessin » du circuit électrique et la position initiale des électrons et traces ;
- la classe CellStructureWireworld qui intègre les variables représentant les éléments de l'espace ;
- la classe CellModelWireworld qui contient la fonction de transition de l'automate.

Cet automate présente une fonction de transition très semblable à celle du Jeu de la Vie. Comme Wireworld repose sur un automate cellulaire à quatre états, la fonction de transition est plus complexe que celle du Jeu de la Vie qui repose sur un automate binaire. Voici comment nous l'avons implémentée dans la méthode « Change () » de la classe CellModelWireworld (Algorithme IV.4):

```
public void change() {
    Int laSomme = 0;
    WireWorldStructure leNewState = new WireWorldStructure(this._status.get_idx(),
this._status.get_idy());
    for (CellModel mdl : this.voisins) {
        ** On compte parmi les cellules voisines celles qui présentent un électron **
        if (Integer.parseInt(((WireWorldStructure) mdl.get_status()).get_status(),10)== 3) {
            laSomme++;
        }
    }
    ** Si la valeur d'état de la cellule courante indique qu'elle est un fil **
    if (this.get_status_value()== 1) {
        ** Si il y a 1 ou 2 électrons présents dans le voisinage **
        if ((laSomme == 1) || (laSomme == 2)) {
            ** La cellule active aura à l'itération suivante une valeur d'état qui la désignera
comme un électron **
            leNewState.set_status("3");
        } else {
            ** Sinon elle présentera à l'itération suivante une valeur d'état qui la décrira
comme un fil **
            leNewState.set_status("1");
        }
    }
    ** Sinon, si la cellule est un électron **
    } else if (this.get_status_value()==3){
        ** Sa valeur d'état à l'itération suivante correspondra à une trace **
        leNewState.set_status("2");
    }
    ** Sinon, si la cellule a une valeur d'état qui la désigne comme une trace **
    } else if (this.get_status_value()== 2){
        ** Sa valeur d'état à l'itération suivante la présentera comme un fil **
        leNewState.set_status("1");
    }
}
```

```

    ** Sinon, si la valeur d'état de la cellule correspond à une cellule vide **
  } else
    ** La valeur d'état de la cellule reste à une valeur qui la décrit comme une maille
    vide **
    leNewState.set_status("0"); // ==> vide
    this._new_status = leNewState;
  }

```

Algorithme IV.4 : Algorithme de l'implémentation de Wireworld.

L'application de la nouvelle valeur d'état à la valeur d'état de chaque cellule n'apparaît pas dans cette méthode, elle est réalisée par la méthode « UpdateAll () » de mise à jour de l'automate à travers une méthode de la classe *Population*.

IV.1.2.3.4. MISE EN ŒUVRE

Le comportement obtenu se révèle fidèle à celui attendu en reproduisant correctement quelques exemples de structures comme des fils simples et des composants simples (diodes, portes logiques, ... Figure IV.7).

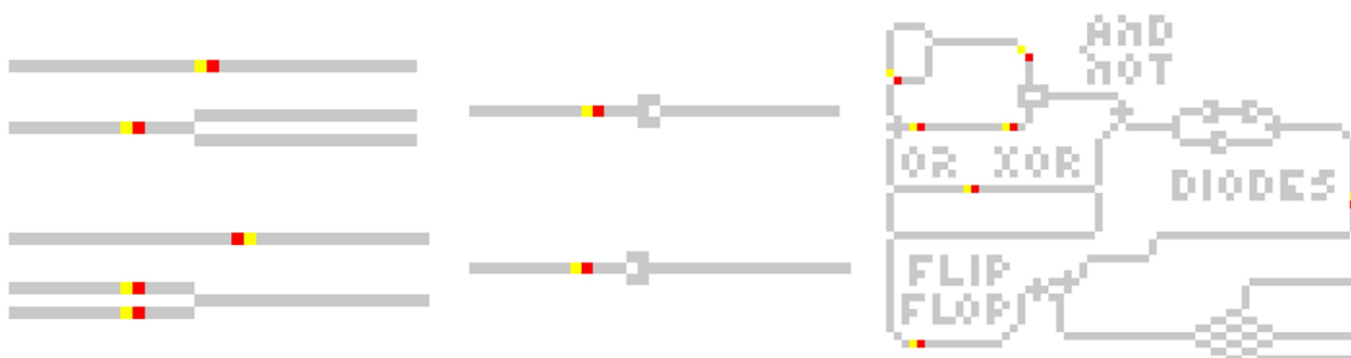


Figure IV.7 : A partir du dessin d'un simple fil, on peut faire des bifurcations dans le circuit (à gauche). De là, des composants peuvent être dessinés, des diodes (au centre) par exemple, que l'on peut agencer pour faire des circuits plus complexes (à droite).

Le résultat d'une simulation de cet automate peut être visualisé en temps réel à l'écran à travers une interface graphique, ou en générant itération par itération un fichier image présentant la configuration de l'automate.

Wireworld nous permet de confirmer la capacité de la plateforme à gérer des automates cellulaires synchrones utilisant plus de deux états et mis à jour selon une fonction de transition déterministe et homogène, avec un maillage régulier de cellules carrées.

Notons qu'en considérant un fil seul, nous avons un automate cellulaire à une dimension. La bonne gestion de ce modèle traduit de fait une bonne gestion des automates cellulaires à une et à deux dimensions.

IV.1.2.4. DAISYWORLD

IV.1.2.4.1. SPECIFICATION

Daisyworld est une construction de l'esprit qui représente un exemple d'écosystème parfait. Il a été imaginé par James Lovelock en 1983 pour illustrer sa théorie de Gaïa.

Daisyworld, dans sa version première, décrit le devenir d'un système stellaire simple constitué d'une planète au relief uniforme et couverte d'une prairie en orbite circulaire autour de son étoile. L'étoile est jeune au début de la simulation et évolue dans le temps pour devenir une vieille étoile. Ce faisant, sa luminosité et sa température varieront. La planète est considérée à distance constante de son étoile. Elle est uniformément couverte d'une prairie herbeuse sur laquelle peuvent apparaître des fleurs qui seront noires ou blanches. La proportion de fleurs blanches et de fleurs noires influe sur la quantité d'énergie captée par la prairie au niveau local et au niveau global. Au niveau local, cela se traduit par la formation de régions plus chaudes ou plus froides, au niveau global, cela se traduit par une fluctuation de la température moyenne. Il existe une rétroaction entre la température et le type de fleurs que l'on peut rencontrer. Une température élevée implique une plus forte probabilité de voir apparaître des fleurs blanches, et inversement une température plus basse favorise les fleurs noires. Les températures extrêmes, trop froide ou trop chaude, sont dans tous les cas incompatibles avec la vie.

À sa naissance, l'étoile éclaire faiblement et ne réchauffe pas suffisamment la planète. En devenant une étoile mature elle devient plus lumineuse et réchauffe la planète jusqu'à permettre à la vie d'émerger. La température de l'atmosphère encore basse est favorable à l'apparition de fleurs noires. Ces dernières capteront plus d'énergie lumineuse et permettront un réchauffement rapide de l'atmosphère, jusqu'à ce que la température devienne assez importante pour permettre l'apparition des fleurs blanches qui à leur tour en reflétant davantage d'énergie induiront une baisse de la température. Pendant ce temps, l'étoile qui continue de vieillir devient toujours plus lumineuse et chauffe toujours plus la planète, si bien que les fleurs blanches ne pourront plus réguler la température, la température augmentera alors au détriment des fleurs noires jusqu'à ce que l'atmosphère devienne si chaude qu'elle en devient incompatible avec la vie.

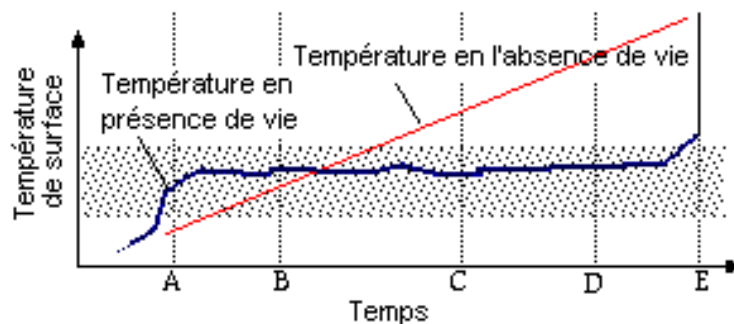


Figure IV.8 : Exemple de résultat classique du modèle Daisyworld ([133]).

Le but de ce modèle est d'observer comment les êtres vivants influencent leur environnement pour le rendre toujours propice à la vie et comment d'une influence locale émerge une influence globale. En effet, une température basse provoque un boom de la

population de fleurs noires qui s'ensuit d'une hausse de la température favorisant le retour des fleurs blanches, et ainsi de suite.

Le modèle original s'attache à suivre la température à la surface de la planète en fonction de l'âge de l'étoile, et donc de sa luminosité, et les proportions des différentes fleurs en fonction de la température. La température suit initialement la luminosité de l'étoile. Dès l'émergence de la population de fleurs noires, la température augmente rapidement, beaucoup plus que la luminosité de l'étoile, jusqu'à atteindre un équilibre lié à la proportion des fleurs noires et blanches. Lorsque la luminosité de l'étoile justifie la température, les fleurs noires déclinent au profit des fleurs blanches cependant que la température reste maintenue à l'équilibre. La luminosité de l'étoile vieillissante continue de réchauffer de plus en plus fort l'atmosphère de la planète ce qui amène à la disparition des fleurs noires, il n'y a alors plus de régulation possible, l'atmosphère voit alors sa température augmenter en suivant la luminosité de l'étoile. Lorsque la température dépasse le seuil de compatibilité avec la vie, les fleurs blanches disparaissent rapidement ainsi que la rétroaction qu'elles exercent sur la température. L'atmosphère se réchauffe rapidement et en suivant l'augmentation de la luminosité de l'étoile (Figure IV.8).

Les versions plus récentes de Daisyworld font intervenir plusieurs dizaines d'espèces végétales et des espèces animales prédatrices qui participent à la régulation de l'écosystème. Plusieurs autres paramètres entrent également en jeu, la planète n'est plus uniforme mais présente un relief plus élaboré, la composition de l'atmosphère peut varier et la physique est prise en compte (rotondité, écliptique, précession, ...).

La rétroaction de l'environnement sur les individus et la modification de l'environnement par les individus sont les éléments de ce modèle que nous avons cherché à reproduire. Nous avons formalisé ce modèle de deux façons. Dans la première, nous nous focalisons sur le phénomène d'autorégulation qui apparaît par l'interaction des populations de fleurs blanches et noires à travers leur influence respective sur l'environnement. La seconde consiste en une traduction du modèle de Daisyworld sous la forme d'un modèle individu-centré dont les individus prennent place dans les mailles d'un automate cellulaire.

IV.1.2.4.2. SIMPLIFICATION DE DAISYWORLD POUR L'ETUDE DE L'HOMEOSTASIE EN SYSTEME ISOLE

Dans cette première version de Daisyworld que nous avons manipulé, nous nous sommes intéressés à la phase de régulation de la température en fonction des proportions de fleurs blanches et noires en négligeant le rôle perturbateur de la luminosité croissante de l'étoile (Figure IV.9)

Nous nous sommes orientés vers un automate cellulaire à trois états qui évolue selon une fonction de transition probabiliste prenant en considération un voisinage de Moore et un voisinage global. La matrice est constituée de mailles régulières, l'application des règles est homogène et synchrone. Le calcul de la température en fonction de la proportion de fleurs de chaque type correspond à la génération d'un champ de force au sein de l'automate par les individus présents dans l'automate. La capacité des fleurs à réagir en fonction de la

température correspond ainsi à la possibilité pour chaque individu de réagir à la présence et à la valeur de ce champ.

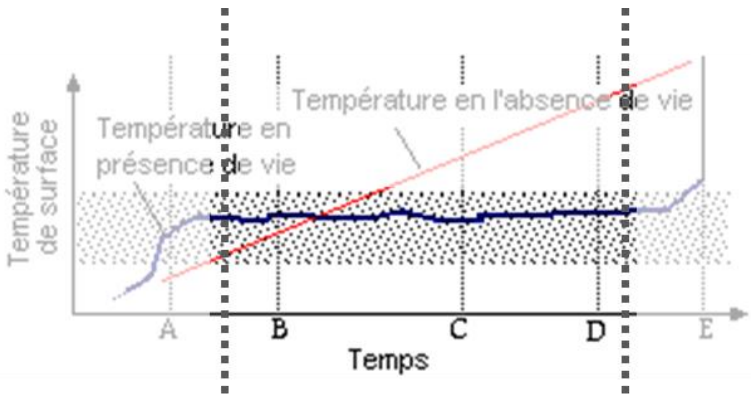


Figure IV.9 : Pour réaliser notre implémentation de Daisyworld, nous avons focalisé sur la phase de régulation de la température par la rétroaction que les populations florales exercent sur elle.

Les trois états pris en compte sont :

- Fleur Blanche ;
- Herbe ;
- Fleur Noire

Les règles de transitions ont été définies pour décrire le réchauffement lié à la forte proportion de fleurs noires, et le rafraichissement induit par une grande proportion de fleurs blanches, ainsi que la probabilité d’obtenir l’un ou l’autre type de fleurs en fonction de la température. De plus, si un seul type de fleurs est présent dans la maille courante et son voisinage, il est impossible que la maille courante présente l’autre type de fleur à l’itération suivante : par exemple, si la maille courante présente une fleur noire, et qu’il n’y a que des fleurs noires dans le voisinage de Moore alors la probabilité d’avoir une fleur blanche dans la maille courante à l’itération suivante est réduite à 0 (Tableau IV.2).

Proportion de fleurs blanches dans le voisinage de Moore	1	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	0,9	0%	12%	23%	35%	46%	58%	69%	81%	90%	91%	93%	94%	95%	96%	98%	100%
	0,8	0%	11%	21%	32%	43%	53%	64%	74%	80%	83%	85%	88%	90%	93%	95%	100%
	0,7	0%	10%	19%	29%	39%	48%	58%	68%	70%	74%	78%	81%	85%	89%	93%	100%
	0,6	0%	9%	18%	26%	35%	44%	53%	61%	60%	65%	70%	75%	80%	85%	90%	100%
	0,5	0%	8%	16%	23%	31%	39%	47%	55%	50%	56%	63%	69%	75%	81%	88%	100%
	0,4	0%	7%	14%	21%	28%	34%	41%	48%	40%	48%	55%	63%	70%	78%	85%	100%
	0,3	0%	6%	12%	18%	24%	30%	36%	42%	30%	39%	48%	56%	65%	74%	83%	100%
	0,2	0%	5%	10%	15%	20%	25%	30%	35%	20%	30%	40%	50%	60%	70%	80%	100%
	0,1	0%	4%	8%	12%	16%	20%	24%	28%	10%	21%	33%	44%	55%	66%	78%	100%
	0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	5	7,5	10	12,5	15	17,5	20	22,5	25	27,5	30	32,5	35	37,5	40	42,5	45
Température																	

Tableau IV.2 : Probabilité d’avoir une fleur blanche dans la cellule active en fonction de la température et de la proportion de fleurs blanches déjà présentes dans son voisinage de Moore.

Pour la construction des règles de cet automate, les paramètres ont été fixés tels que suit :

- 1% de fleurs blanches induit une baisse de 0,25 °C de la température globale ;
- 1% de fleurs noires induit un réchauffement global de 0,25 °C ;
- La température moyenne est fixée à 25 °C ;
- 20% de la surface totale de l'écosystème sera toujours occupé par de l'herbe (neutre vis-à-vis de la température). Ainsi les règles s'écrivent :
- La température est calculée à chaque itération en fonction de la proportion de chaque type de fleurs :
 - o S'il n'y a que des fleurs noires, soit 80% de la surface, la température globale montera à 45 °C ($25 + 80 * 0,25 = 45$) l'écart à la température moyenne est alors de +20 C ;
 - o S'il n'y a que des fleurs blanches, soit 80% de la surface la température globale baissera à 5 °C ($25 - 80 * 0,25 = 5$) l'écart à la température moyenne est alors de +20 C.
- Le type de végétation est choisi aléatoirement, la probabilité d'avoir une fleur blanche ou noire est calculée en fonction de l'écart entre la température calculée, et la température moyenne ; avec une modulation par la présence des types de fleurs dans le voisinage de Moore comme présenté dans le Tableau IV.2.

IV.1.2.4.3. ANALYSE

Les données des cellules sont stockées dans les attributs de la classe *CellStructureGaia* dérivée de la classe *CellStructure* (Figure IV.10).

L'algorithme décrivant la fonction de transition de cet automate cellulaire et les différentes variables locales sont employées dans la conception de la classe *CellModelGaia* (Figure IV.10) dérivée de la classe *CellModel*.

L'architecture de la plateforme permet d'envisager facilement de décrire les deux versions de Daisyworld de façon similaire. La différence se situe dans le contenu des attributs et des méthodes des classes, et principalement dans les détails de la fonction de transition.

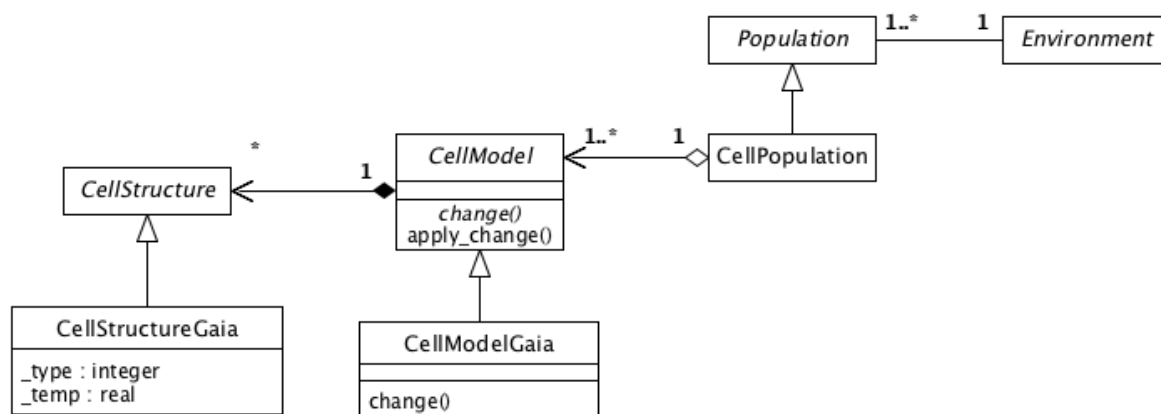


Figure IV.10 : Diagramme de classe de l'implémentation de Daisyworld.

La gestion de l'augmentation de la température corrélée à l'augmentation de la luminosité de l'étoile est confiée à la classe *Environnement*. La température locale qui apparaît en

attribut de la classe CellStructureGaia est manipulée par la méthode « Change () » de la classe CellModelGaia. Elle est calculée en fonction de la température de l'environnement et des types de fleurs présents dans le voisinage. La température de l'environnement est en retour calculée dans le temps en fonction de la luminosité de l'étoile et en fonction des températures locales.

A. VERSION SIMPLIFIEE DE DAISYWORLD POUR L'ETUDE DE L'HOMÉOSTASIE

Pour élaborer ce modèle, nous avons défini 8 données par cellule :

- $_state_i \in \{\text{Fleur blanche ; Fleur noire ; Herbe}\}$ pour représenter l'état courant de la cellule i ;
- $_futureState_i \in \{\text{Fleur blanche ; Fleur noire ; Herbe}\}$ pour représenter l'état de la cellule i à l'itération suivante ;
- $N_i =$ Ensemble de cellules comprenant la cellule i et les cellules comprises dans le voisinage de Moore de la cellule i (soit une liste de 9 cellules avec un voisinage de Moore de rayon 1 ; la cellule i et ses 8 cellules voisines)
- ProbaFlore tableau $[0..10] [0..16]$ réels (Tableau IV.2)
- $Temp \in \mathbb{R}$, la température de l'environnement ;
- $Taille \in \mathbb{N}$, le nombre total de cellules ;
- $Blanches \in \mathbb{N}$, le nombre de fleurs blanches dans l'espace ;
- $Noires \in \mathbb{N}$, le nombre de fleurs noires dans l'espace.

Ainsi que 6 variables locales :

- $Végétation \in [0, 1]$ une variable aléatoire entre 0 et 1 ;
- $Fleur \in [0, 1]$ une variable aléatoire entre 0 et 1 ;
- $nbB \in \mathbb{N}$, le nombre de fleurs blanches dans le voisinage d'une maille ;
- $ratioB \in [0, 1]$, la proportion de fleurs blanches dans le voisinage d'un maille ;
- $Ligne \in \mathbb{N}$, le numéro de ligne du tableau contenant l'information de probabilité voulue ;
- $Colonne \in \mathbb{N}$, le numéro de colonne du tableau contenant l'information de probabilité voulue.

L'algorithme de mise à jour de cet automate s'écrit alors (Algorithme IV.5):

```

Blanches ← 0
Noires ← 0
Pour chaque cellule  $i$  faire
|    $nbB \leftarrow 0$ 
|    $Végétation \leftarrow$  une valeur tirée aléatoirement entre 0 et 1
|    $Fleur \leftarrow$  une valeur tirée aléatoirement entre 0 et 1
|   ** On regarde ce qu'il y a dans la cellule  $i$  et on parcourt son voisinage **
|   Pour chaque cellule  $n$  de  $N_i$  faire
|   |   ** On compte les fleurs blanches **
|   |   Si  $Etat_n = \text{Fleur blanche}$  alors
|   |   |    $nbB \leftarrow nbB + 1$ 
|   |   Fin si
|   Fin pour

```

```

/      ** En divisant ce nombre de fleurs blanches dans le voisinage par le nombre cellules prise
en compte, dans ce cas la cellule i et son voisinage de Moore, soit 9 cellules au total, on calcule
la proportion de fleurs blanches dans le voisinage de la cellule i **
|      ratioB ← nbB/9
|      Ligne ← ⌊ ratioB ⌋
|      Colonne ← ⌊ (Temp - 5) / 2,5 ⌋
/      ** On réserve 20% de l'espace qui ne portent pas de fleurs **
|      Si Vegetation ≤ 0,2 alors
|      |      _futureStatei ← Herbe
/      ** Pour les autres cellules, on évalue si la valeur d'état correspondra à une fleur noire ou
blanche à l'itération suivante**
|      Sinon
|      |      Si Fleur ≤ ProbaFlore [Ligne][Colonne] alors
|      |      |      _futureStatei ← Fleur noire
|      |      |      Noires ← Noires + 1
|      |      Sinon
|      |      |      _futureStatei ← Fleur blanche
|      |      |      Blanches ← Blanches + 1
|      |      Fin si
|      Fin si
Fin pour
Temp ← 25 + (0,25 * Noires/Taille) – (0,25 * Blanches/Taille)

```

Algorithme IV.5 : Algorithme de principe de la version simplifiée de Daisyworld.

B. VERSION CLASSIQUE DE DAISYWORLD

Pour décrire ce modèle, nous avons utilisé 4 données par cellule :

- $_state_i \in \{\text{Fleur blanche ; Fleur noire ; Herbe}\}$ pour représenter l'état courant de la cellule i ;
- $_futureState_i \in \{\text{Fleur blanche ; Fleur noire ; Herbe}\}$ pour représenter l'état de la cellule i à l'itération suivante ;
- N_i = Liste de cellules comprenant la cellule i et les cellules comprises dans le voisinage de Moore de la cellule i (soit une liste de 9 cellules avec un voisinage de Moore de rayon 1 ; la cellule i et ses 8 cellules voisines) ;
- $_localTemp_i \in \mathbb{R}$ qui stocke la température de la cellule i calculée en fonction du type de fleur qu'elle présente et des fleurs présentes dans son voisinage.

Nous avons également utilisé 6 variables locales :

- $nbVoisinsNoirs \in \mathbb{N}$, le nombre de cellules de N_i dont la valeur d'état désigne une fleur noire ;
- $nbVoisinsBlancs \in \mathbb{N}$, le nombre de cellules de N_i dont la valeur d'état désigne une fleur blanche ;
- $temp \in \mathbb{R}$, la température de la planète
- $aléa \in [0, 1]$, une variable aléatoire entre 0 et 1
- $Végétation \in [0, 1]$, une variable aléatoire entre 0 et 1
- $blancOuNoir \in [0, 1]$, une variable aléatoire entre 0 et 1

L'algorithme de mise à jour s'écrit dans ce cas (Algorithme IV.6) :

Pour chaque cellule i **faire**

```

|   aléa ← une valeur tirée aléatoirement entre 0 et 1
|   Végétation ← une valeur tirée aléatoirement entre 0 et 1
|   blancOuNoir ← une valeur tirée aléatoirement entre 0 et 1
/   ** On examine la valeur d'état de la cellule  $i$  et de celles dans son voisinage de Moore **
|   Pour chaque cellule  $n$  de  $N_i$  faire
|   /   ** On comptabilise les cellules dont la valeur d'état désigne une fleur noire **
|   |   Si  $\text{state}_n = \text{Fleur noire}$  alors
|   |   |    $\text{nbVoisinsNoirs} \leftarrow \text{nbVoisinsNoirs} + 1$ 
|   |   /   ** On comptabilise les cellules dont la valeur d'état désigne une fleur blanche **
|   |   Sinon si  $\text{state}_n = \text{Fleur blanche}$  alors
|   |   |    $\text{nbVoisinsBlancs} \leftarrow \text{nbVoisinsBlancs} + 1$ 
|   |   Fin si
|   /   ** On stocke les températures locales pour calculer la température moyenne de la
|   cellule  $i$  **
|   |    $\text{temp} \leftarrow \text{temp} + \text{localTemp}_n$ 
|   Fin pour
|   Selon  $\text{state}_i$ 
|   /   /   ** Si la cellule  $i$  présente une valeur d'état qui la désigne comme une Fleur noire **
|   |   Cas Fleur noire :
|   /   /   /   ** La température de la cellule  $i$  est calculée en moyennant les températures
|   des cellules du voisinage  $N_i$  de la cellule  $i$  **
|   |   |    $\text{localTemp}_i \leftarrow (\text{temp} + (0,75 * \text{localTemp}_i)) / 9$ 
|   /   /   /   ** Une valeur d'état indiquant une Fleur noire dans la cellule  $i$  induit un
|   réchauffement local qui est répercuté sur la température locale de toutes les cellules  $n$  de  $N_i$  **
|   |   |   Pour chaque cellule  $n$  de  $N_i$  faire
|   |   |   |    $\text{localTemp}_n \leftarrow \text{localTemp}_n + 3$ 
|   |   |   Fin pour
|   /   /   /   ** Si la température locale est trop basse, aucune fleur ne peut pousser **
|   |   |   Si  $\text{localTemp}_i < 5$  alors
|   |   |   |    $\text{futureState}_i \leftarrow \text{Herbe}$ 
|   |   |   Sinon
|   /   /   /   ** Une température trop élevée n'autorise que la naissance d'une Fleur
|   blanche **
|   |   |   |   Si  $\text{localTemp}_i \geq 40$  et  $\text{aléa} \leq 0,1$  alors
|   |   |   |   |    $\text{futureState}_i \leftarrow \text{Fleur blanche}$ 
|   |   |   |   Fin si
|   |   |   Fin si
|   /   /   ** Si la cellule  $i$  présente une valeur d'état qui la désigne comme une Fleur blanche
|   **
|   |   Cas Fleur Blanche
|   /   /   /   ** Une valeur d'état indiquant une Fleur blanche dans la cellule  $i$  induit un
|   refroidissement local qui est répercuté sur la température locale de toutes les cellules  $n$  de  $N_i$  **
|   |   |    $\text{localTemp}_i \leftarrow (\text{temp} + (0,25 * \text{localTemp}_i)) / 9$ 
|   |   |   Pour chaque cellule  $n$  de  $N_i$  faire
|   |   |   |    $\text{localTemp}_n \leftarrow \text{localTemp}_n - 3$ 
|   |   |   Fin pour
|   /   /   /   ** Si la température locale est trop élevée, aucune fleur ne peut pousser **
|   |   |   Si  $\text{localTemp}_i > 40$ 
|   |   |   |    $\text{futureState}_i \leftarrow \text{Herbe}$ 

```

Algorithme IV.6 : Algorithme de principe de Daisyworld

IV.1.2.4.4. IMPLEMENTATION DE DAISYWORLD

L'implémentation passe par la conception des classes correspondant à ce modèle. Elles sont au nombre de trois :

- la classe `fileReader` qui permet de spécifier la configuration qui va être employée pour la simulation, la répartition et la disposition des différentes populations florales ;
- la classe `CellStructureGaia` qui intègre les variables représentant les cellules de l'espace ;
- la classe `CellModelGaia` qui contient la fonction de transition de l'automate.

Que l'on s'intéresse à la version classique ou à la version simplifiée de Daisyworld ne change pas les classes qui doivent être préparées. La différence d'approche entre ces deux versions de ce modèle tient au contenu de la méthode « `Change ()` » de la classe `CellModelGaia`.

L'application de la nouvelle valeur d'état à la valeur d'état de chaque cellule n'apparaît pas dans la méthode « `Change ()` », elle est déléguée à la mise à jour de l'automate à travers une méthode de la classe `Population`. Comme Daisyworld n'est pas un modèle d'automate cellulaire dans sa définition, nous l'avons exécuté avec les différents processus de mise à jour, synchrone (« `UpdateAll ()` » et « `UpdateAllRand ()` ») et asynchrone (« `UpdateAllAsynchrone` » et « `UpdateAllAsynchroneRand ()` »).

IV.1.2.4.5. MISE EN ŒUVRE DE DAISYWORLD

Les tests réalisés avec ce modèle ont été menés avec des matrices de tailles différentes, de 5x5 à 1 000x 1 000. La version simplifiée de Daisyworld nous permet de suivre l'évolution de la température des proportions des populations florales (Figure IV.11).

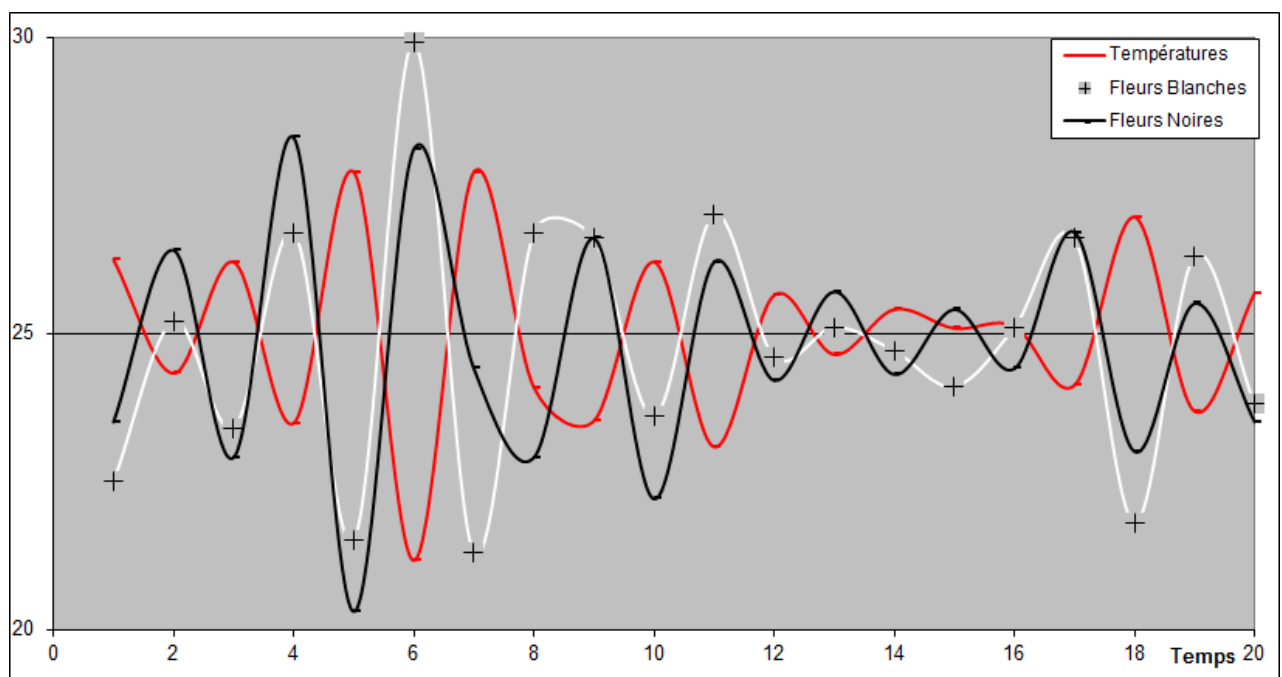


Figure IV.11 : Exemple de suivi de la température en fonction des proportions de chaque type de fleurs dans l'environnement en utilisant une mise à jour synchrone de l'automate cellulaire.

L'implémentation classique de Daisyworld reproduit les courbes que l'on connaît dans la littérature. La température s'accroît d'abord linéairement avec la luminosité de l'étoile jusqu'à ce qu'elle devienne propice à la vie. Une brusque efflorescence noire la fait alors augmenter rapidement. La température devient favorable aux fleurs blanches qui contribuent à refroidir la planète alors que l'étoile continue à réchauffer de plus en plus le système. Les courbes des deux populations se croisent et la température de la planète reste stable pendant un temps. Lorsque l'étoile chauffe trop fort et que les fleurs noires ont disparu, la température recommence à augmenter fortement jusqu'à redevenir incompatible avec la vie ce qui conduit à l'extinction des fleurs blanches.

Cette implémentation est sensible à la taille de l'espace, à la différence de la version simplifiée, et l'utilisation d'une matrice trop petite, de dimensions 80x80 ou moins, se traduit par des oscillations autour d'un comportement moyen, ce que nous supposons lié à une forte influence locale des règles stochastiques qui est « lissée » avec l'augmentation de la taille de l'espace. Ces oscillations sont d'autant plus importantes que la matrice est petite, et c'est un comportement normal car les règles de fonctionnement du modèle impactent le voisinage considéré lors de la mise à jour d'une cellule. Si le voisinage représente une proportion importante de la matrice, l'effet est fortement marqué. Le même voisinage représente une proportion plus restreinte de l'espace, et l'effet de la mise à jour sur les mailles du voisinage est ainsi lissé par la surface. Plus la matrice est grande, moins l'effet est important. Grossièrement, on peut considérer qu'un espace plus large a une plus grande inertie, ce qui est cohérent avec le système modélisé.



Figure IV.12 : Trois exemples d'écosystèmes produits. À gauche, un équilibre est installé entre les populations de fleurs noires et de fleurs blanches. Les deux autres exemples correspondent à des dérives extrêmes qui ont conduit à la disparition de l'une ou l'autre population florale.

Le déroulement de la simulation de ce modèle sur la plateforme peut être visualisé en temps réel à l'écran à travers une interface graphique qui peut présenter l'image de la population à chaque instant et les graphiques de suivi des populations et de la température dans le temps, ou en générant des fichiers correspondant itération par itération à un fichier image présentant la configuration de l'automate et à la fin de la simulation aux listes des valeurs des populations et de la température à chaque instant.

IV.1.2.5. LIEN AVEC D'AUTRES TRAVAUX DE L'EQUIPE

Fazia Aiboud a développé dans sa thèse (Aiboud, 2013 [134]) des méthodes permettant d'identifier la fonction de transition d'un automate cellulaire pour obtenir une forme voulue à partir d'une condition initiale connue, ce qui peut par exemple s'appliquer pour étudier des processus de morphogénèse.

Par exemple, dans le cas d'un automate cellulaire binaire, déterministe, synchrone, homogène et exploitant comme système de voisinage la cellule courante et le voisinage de Von Neumann, on détermine les règles de transition en procédant pour chaque cellule à t à la lecture de l'état ses voisines à t .

Codage décimal	Règle de transition	
	Vecteur de valeurs binaires à t	État à $t+1$
0	00000	1
1	00001	0
2	00010	0
3	00011	1
4	00100	1
...
29	11101	0
30	11110	1
31	11111	0

Tableau IV.3 : Exemple de fonction de transition d'un automate cellulaire binaire.

On obtient alors un vecteur de valeurs binaires qui se traduit par une valeur décimale. Par exemple, la suite de valeurs 01001 vaut 9. Dans le cas pris en exemple, le nombre de combinaisons dans le voisinage est 2^5 , soit 32 valeurs qui sont représentées par le vecteur 00000 qui vaut 0 au vecteur 11111 qui vaut 31. Pour chacun de ces vecteurs on associe la valeur d'état de la cellule à $t+1$. La fonction de transition alors obtenue se présente sous la forme d'un vecteur de valeurs binaires qui correspondent respectivement pour chaque valeur de voisinage à la valeur de l'état à $t+1$ de la cellule courante (Tableau IV.3).

Au sein de la plateforme, on stocke alors ce vecteur solution en indexant chacune de ces valeurs. On évalue pour chaque cellule la valeur de son voisinage dans la fonction « Change () », et on utilise cette valeur comme index pour récupérer la valeur de l'état $t+1$.

En définissant comme variable :

- Règles tableau [0..255] booléens
- N_i l'ensemble des cellules voisines de la cellule i
- Voisinage, la valeur décimale codée par le vecteur d'état des cellules de N_i

La fonction de transition peut alors exploitée en suivant l'Algorithme IV.1 :

Début

```

|   Pour chaque cellule  $i$  faire
|   |    $\_futureState_i \leftarrow$  Règle [Voisinage]
|   Fin pour
|   Pour chaque cellule  $i$  faire
```

```

|         |         _statei ← _futureStatei
|         |         Fin pour
Fin

```

Algorithme IV.1 : Algorithme d'utilisation d'une fonction de transition codée sous forme d'un vecteur de valeurs binaires.

Avec cette notation, la fonction de transition du Jeu de la Vie est représentée par le vecteur suivant, chaque valeur représentant une règle de transition :

```
0;0;0;1;0;1;1;1;0;1;1;1;1;1;1;0;0;1;1;1;1;1;1;0;1;1;1;0;1;0;0;0;
;0;1;1;1;1;1;1;0;1;1;1;0;1;0;0;0;1;1;1;0;1;0;0;0;1;0;0;0;0;0;0;
0;0;1;1;1;1;1;1;0;1;1;1;0;1;0;0;0;1;1;1;0;1;0;0;0;1;0;0;0;0;0;0;
;0;1;1;1;0;1;0;0;0;1;0;0;0;0;0;0;0;1;0;0;0;0;0;0;0;0;0;0;0;0;0;
0;0;0;1;1;1;1;1;1;0;1;1;1;0;1;0;0;0;1;1;1;0;1;0;0;0;1;0;0;0;0;0;
;0;0;1;1;1;0;1;0;0;0;1;0;0;0;0;0;0;0;1;0;0;0;0;0;0;0;0;0;0;0;0;
0;0;0;1;1;1;0;1;0;0;0;1;0;0;0;0;0;0;0;1;0;0;0;0;0;0;0;0;0;0;0;
;0;0;0;1;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;
0;0;0;0;
```

IV.1.3. REMARQUES SUR LES TEMPS DE CALCULS

Les durées des simulations augmentent très vite avec la taille de la matrice (Figure IV.13) et avec la taille du voisinage (Figure IV.14). Avec une matrice de 100x100 avec un voisinage de Moore R1, la simulation ne prend qu'un peu plus d'une minute, avec une matrice de 500x500 et le même voisinage, la simulation dure environ 25 minutes. Si l'on exploite ces deux matrices en utilisant un voisinage de Moore de rayon 3, la simulation durera respectivement, deux minutes trente environ et 75 à 80 minutes.

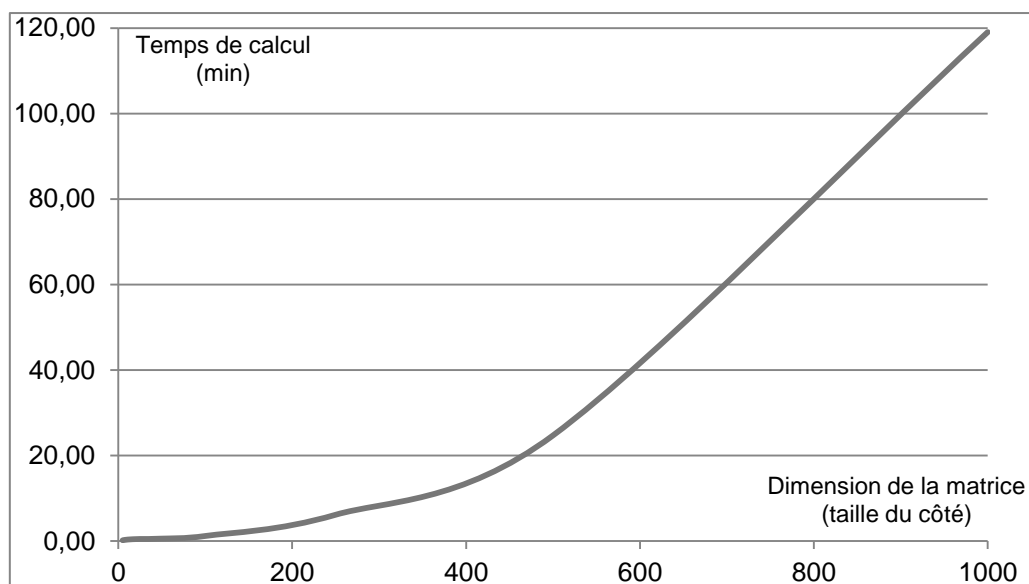


Figure IV.13 : Temps de calcul en fonction de la taille de la matrice considérée

Le code de la plateforme est actuellement parallélisé pour gérer les répliques. Chaque nouvelle réplique démarre un thread. On peut ainsi lancer en parallèle plusieurs simulations, soit avec les mêmes paramètres, soit avec des jeux de paramètres différents. Un élément important du coût en temps de calcul est aussi que nous avons pris le parti de mettre à jour tout l'automate à chaque itération, même si une maille est vide. Cette hypothèse nous permet d'explorer les différents modèles avec notre plateforme sans apporter de modification ni aux modèles, ni à la gestion de l'espace, et avec le modèle de cellule biologique cela se traduit par la possibilité de manipuler indifféremment l'hypothèse d'une culture en milieu liquide, agité, ou en milieu solide. Si cet aspect se paye en termes de coût de calcul, il apporte une généralité à la plateforme.

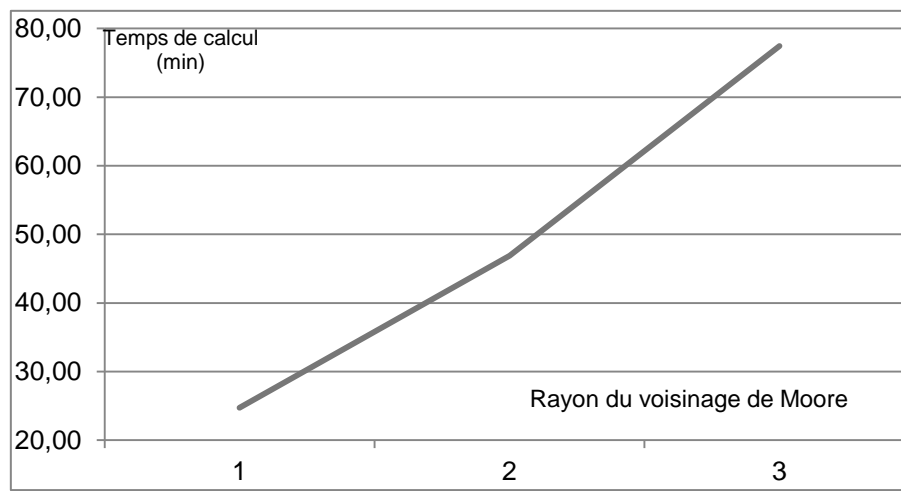


Figure IV.14 : Temps de calcul en fonction du rayon du voisinage de Moore utilisé

Si l'on veut optimiser cet aspect, il faut apporter les changements dans la classe CellPopulation. Ce changement devra également être répercuté dans le FileReader et dans le CellModel. Le premier car c'est lors de son appel que se fait l'initialisation de l'automate et donc la création du voisinage, et par conséquent, que se décide sa topologie. Le CellModel doit être modifié pour intégrer la façon de consulter ce voisinage. En effet, au lieu d'instancier une matrice complète avec ses sites libres ou occupés, une telle optimisation passerait par l'instanciation « ciblée » des sites occupés seulement. Par conséquent, au lieu de surveiller les sites libres de son voisinage, une cellule devrait surveiller le nombre de voisins qu'elle connaît et demander à en instancier un nouveau si elle en a moins que ce qui est défini dans la topologie de l'automate. Notons que dans ce cas, le modèle connaît le maillage alors qu'il s'agit là d'une hypothèse que nous avons écartée pour l'instant. Notre hypothèse est que l'espace a une structure que les individus ne peuvent pas connaître, les individus ne savent que consulter l'espace qui leur est accessible, et ce, à travers le système de voisinage utilisé dont la structure est inconnue du modèle. Ceci permet de moins introduire d'hypothèses de modélisation dans la règle de transition de l'automate cellulaire.

Rajoutons que ce genre d'optimisation, s'il permet d'accélérer la simulation tant que peu de sites sont occupés, ne change rien dès que la matrice est remplie.

IV.1.4. CONCLUSION

La plateforme s'est révélée fonctionnelle et capable de faire tourner des automates cellulaires aux propriétés variées :

- synchrone ou asynchrone ;
- déterministe ou probabiliste ;
- nombre d'états varié ;
- voisinages variés ;
- à 1 ou 2 dimensions.

Quel que soit le modèle que nous avons testé sur la plateforme proposée pour en valider les caractéristiques, il ne faut toujours fournir que trois classes, quelle que soit la complexité du modèle, la classe dérivée de *CellStructure*, la classe dérivée de *CellModel* et un *fileReader*. Si ce dernier a pour fonction de permettre de fournir des conditions initiales voulues pour réaliser une simulation, on peut remplacer son contenu pour permettre la génération aléatoire d'une instance, ou pour permettre à l'utilisateur de choisir entre un cas particulier et un cas généré aléatoirement.

La classe dérivée de *CellStructure* contient tous les attributs nécessaires à la définition et à la caractérisation d'un individu d'un modèle individu-centré, et en l'occurrence, à la description d'une maille d'un automate cellulaire.

La classe dérivée de *CellModel* présente les règles d'interactions, regroupées dans la méthode « *Change ()* ». Dans le cadre d'un automate cellulaire, il s'agit de fournir dans cette méthode les règles de transition. Il est difficile de parler de fonction de transition dans ce cas dans la mesure où la mise à jour des cellules et l'application d'une nouvelle valeur d'état calculée à chaque cellule est déléguée à la classe *Population*.

Ces propriétés sont résumées dans le Tableau IV.4.

	Game Of Life	Wireworld	Daisyworld	FONCTIONNALITES VALIDEES
DIMENSIONS	2	1, 2	2	1D ou 2D
ÉTATS	2	3 (ou 4)	3	2 états et plus
VOISINAGE	Moore	Moore	RMoore (*)	RMoore (*)
INTERACTIONS CELLULES- POPULATION	—	—	Cellules ↔ Population	Bidirectionnelles entre les cellules et la population
MISE A JOUR	Synchrone	Synchrone	Synchrone ou asynchrone	Synchrone et asynchrone
REGLES	Déterministes	Déterministes	Probabilistes	Déterministes et probabilistes

Tableau IV.4 : Caractéristiques des automates cellulaires pris en compte dans l'élaboration de la plateforme. Le terme de RMoore (*) indique l'exploitation d'un voisinage de Moore étendu, de rayon supérieur à 1, jusqu'à un rayon infini qui désigne l'intégralité du maillage.

Chaque automate a pu être implémenté sans changer les éléments de la plateforme, la plateforme s'est ainsi démontrée générique, au sens où elle peut faire tourner différents

types d'automates cellulaires du moment que les règles fonctionnelles de l'automate sont passées à la plateforme selon le format requis.

() Remarque : Lorsque le rayon du voisinage de Moore utilisé est infini, cela correspond à l'exploitation dans le cadre du voisinage du maillage complet de l'automate cellulaire pour la mise à jour des cellules. Cela correspond à la génération par les individus d'un champ, qui est pris en compte par l'ensemble de la population.*

L'environnement décrit dans la classe éponyme peut avoir une influence sur le comportement des individus du modèle, et ceux-ci peuvent en retour lui imposer des modifications. Ceci nous permet de modéliser des boucles de rétroaction, qui se traduiront dans un modèle de population cellulaire comme le fait que les cellules sont dépendantes de leur environnement, mais qu'elles sont capables de le modifier.

IV.2. MODELE DE CELLULE

Nous avons vu dans la description des systèmes biologiques que les cellules fonctionnent indépendamment les unes des autres en exploitant leur potentiel qui est porté par leur information génétique. L'exploitation de ce potentiel résulte de l'exploitation du patrimoine génétique à l'aide de la machinerie cellulaire et de l'interaction entre le génome, l'environnement et la machinerie cellulaire. Nous avons vu dans l'état de l'art que les modèles individu-centrés sont les plus indiqués pour modéliser ce fonctionnement car ils présentent des individus qui disposent de leurs propres règles de fonctionnement.

Cela n'empêche pas les individus de coopérer, si la coopération fait partie de leur règle de fonctionnement, ce qui correspond à ce qui peut être observé dans le vivant, les organismes d'une même espèce ou d'espèces différentes peuvent coopérer si cela leur est favorable, et les cellules d'un organisme coopèrent pour assurer son bon fonctionnement.

C'est donc vers l'élaboration d'un modèle individu-centré que nous nous sommes orientés. Il faut alors définir les règles de fonctionnement du modèle de cellule. Elles constitueront la fonction `Change` de la classe `CellBiologyModel` dérivée de `CellModel`. Les informations du patrimoine génétique, qui caractérisent l'individu par rapport à l'espèce à laquelle il appartient sont les informations qui permettent l'identification d'un individu, ce sont donc les attributs de l'individu. Ils seront donc présents dans les attributs de la classe `CellBiologyStructure` dérivée de la classe `CellStructure`.

L'environnement dans lequel se développent les cellules représentera un milieu de culture puisque nous avons précisé que la culture bactérienne constitue actuellement notre cadre expérimental de référence. La classe `Environment` permettra de décrire les propriétés de l'environnement qui auront une importance pour les cellules, c'est-à-dire ses caractéristiques physiques (température, pH ; etc...) et ses caractéristiques chimiques (types de ressources disponibles et quantités de chacune, présence d'éléments nocifs et quantité de chaque, etc...).

IV.2.1. HYPOTHESES DE MODELISATION

IV.2.1.1. FONCTIONNEMENT DE LA CELLULE

« Rien ne se perd, rien ne se crée, tout se transforme » (Lavoisier). Nous avons vu lors de la description des systèmes biologiques qu'une cellule maintient son état de vie en échangeant matière et énergie avec son environnement. L'interruption d'un de ces flux conduisant à l'arrêt des fonctions cellulaires, et par suite, à la mort. Les flux entrants vont de l'environnement ou milieu extracellulaire à l'intérieur de la cellule ou milieu intracellulaire. Les flux sortants sont orientés en sens inverse. Pour simplifier, nous parlerons d'absorption pour les flux entrants, et d'excrétion pour les flux sortants.

À partir de cette constatation, un premier schéma du fonctionnement du modèle a été élaboré (Figure IV.15).

- La cellule absorbe des éléments de son environnement
- Avec ce qui est absorbé la cellule assure ses fonctions
 - o La cellule assure les processus vitaux
 - o La cellule s'occupe de maintenir ses constituants en bon état
- La cellule excrète tous les déchets ou produits qui ne lui sont plus utiles, ou qui lui sont utiles à l'extérieur

Des éléments sont absorbés par la cellule, elle les exploite comme elle en a besoin ou comme elle peut par le biais de ses fonctions cellulaires, puis des éléments sont excrétés. Plus explicitement, en termes de matière, on aura principalement des nutriments en absorption (Figure IV.15-1), et des déchets en excrétion (Figure IV.15-3).

La transformation des éléments absorbés en éléments qui vont être excrétés, l'extraction de l'énergie des nutriments, et le maintien en bon état des constituants cellulaires sont le résultat des fonctions cellulaires (Figure IV.15-2) sous lesquelles se regroupent les éléments du métabolisme, à travers l'anabolisme et le catabolisme. Nous avons défini une fonction cellulaire qui permet d'obtenir à partir des flux entrants la matière qui renouvelle les constituants cellulaires, les déchets qui seront éliminés, et l'état bioénergétique de la cellule. Appliquer cette fonction sur les flux entrants nous donne les valeurs des flux sortants (Figure IV.15-3).

Pour cette fonction, nous avons introduit trois notions : celle de consommation, celle de besoins et celle d'état interne.

La consommation regroupe pour l'instant tout ce qui peut être consommé par la cellule, sans considération pour les types (énergie ou matière) ou les espèces chimiques (sucres, vitamines, ...) des éléments consommés. Cette capacité de consommation est dictée par les possibilités de la cellule qui sont données par la génétique, c'est-à-dire par l'espèce à laquelle elle appartient et par sa spécificité en tant qu'individu. Cette capacité est modulée par l'environnement, la température nous l'avons vu jouant un rôle sur l'efficacité du métabolisme, et par l'état interne de la cellule. Une cellule en pleine possession de ses moyens est alors plus efficace qu'une cellule dont les niveaux énergétiques sont bas ou qu'une cellule dont la machinerie cellulaire est abîmée. La consommation est appelée à être

affinée pour considérer les capacités de consommation propres à chaque cellule pour chaque type et espèce différents de consommables qui peuvent être envisagés (Figure IV.15-2a).

Les besoins sont caractéristiques de la cellule. Ils sont définis par la génétique, et donc par l'espèce à laquelle appartient la cellule, en prenant en compte ses particularités individuelles. Dans les besoins, nous avons pour l'instant un seul niveau qui correspond au minimum vital. Ces besoins pourront évoluer pour correspondre à différents paliers, du minimum vital au nécessaire pour que la cellule soit en pleine santé. Là encore, les besoins considèrent actuellement de façon indifférenciée les types (énergie ou matière) et espèces chimiques (sucres, vitamines, ...) requises pour le fonctionnement cellulaire. Ils sont appelés à être affinés pour prendre en compte les besoins d'une cellule pour autant d'entités que nécessaire (Figure IV. 15-2a).

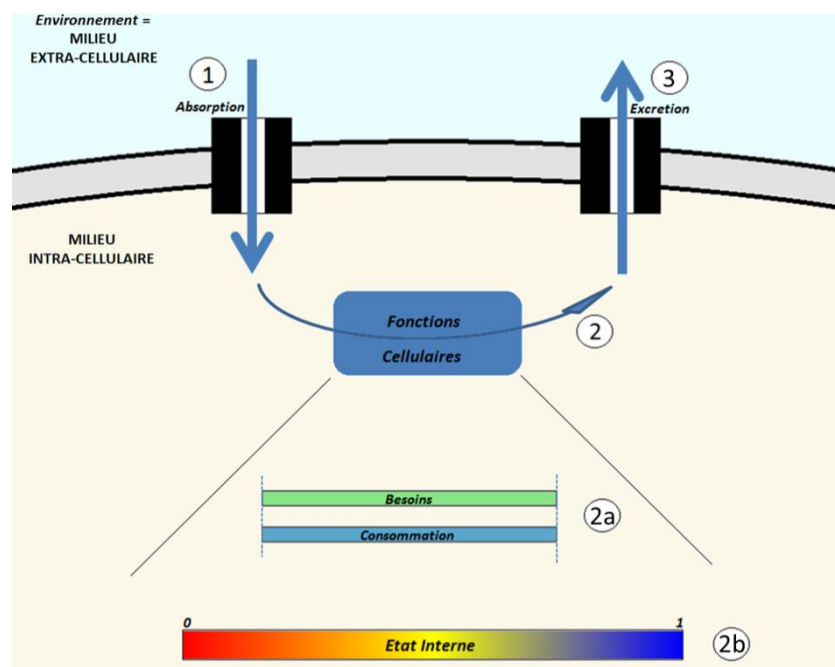


Figure IV.15 : Schéma de principe du fonctionnement d'une cellule.

L'état interne décrit simplement l'état de la cellule. C'est une variable qui indique la bonne santé ou le mauvais état d'une cellule. Sous cette valeur nous regrouperons pour l'instant l'aspect énergétique (potentiels électrochimiques) et l'aspect biochimique (machinerie cellulaire en bon état de marche). Une cellule en pleine forme a un état interne au maximum ou proche. Une carence ou un stress impactera cette valeur et la fera décroître. Des conditions favorables permettront la régénération de l'état interne vers une valeur correcte. Lorsque la valeur de l'état interne devient trop basse, les dommages sont trop importants et deviennent irréversibles et la cellule mourra. Cela sera donc soit que ses potentiels énergétiques sont devenus trop bas pour assurer le maintien des fonctions cellulaires, soit que la machinerie cellulaire sera trop endommagée pour être reconstituée (Figure IV.15-2b).

IV.2.1.2. CYCLE DE VIE DE LA CELLULE

Maintenant que nous disposons d'un modèle qui nous permet de gérer les flux entrants et sortants de la cellule, et donc les échanges entre la cellule et l'environnement, et que nous savons comment faire fonctionner une cellule à partir de ce qui a ainsi été échangé, il nous faut construire autour un modèle permettant de gérer la vie de la cellule, de sa naissance à sa disparition.

Pour cela nous proposons d'itérer sur les échanges et la fonction cellulaire pour chaque cellule de sa naissance et jusqu'à sa disparition. Cette disparition peut survenir soit parce que la cellule n'aura pas pu maintenir son état interne ou qu'une cause extérieure l'aura altéré, et dans ce cas la cellule mourra, soit qu'elle se sera divisée et aura donné naissance à deux nouvelles cellules (Algorithme IV.7). Nous avons choisi de représenter l'état interne sous la forme d'une variable réelle comprise entre 0 et 1 (Figure IV.16).

Ainsi l'algorithme de principe peut se résumer comme suit :

```
Tant que la simulation n'est pas terminée faire
|   Pour chaque cellule faire
|   /   ** On commence par évaluer les valeurs courantes des variables cellulaires **
|   |   Évaluer la quantité de nutriments absorbés
|   |   Calculer la quantité de nutriments consommés et la confronter aux besoins
|   /   ** Si les besoins ne sont pas satisfaits, il y a carence et l'état interne est affecté **
|   |   Mettre à jour l'état interne de la cellule
|   |   Calculer la quantité de composés excrétés
|   /   ** Un état interne trop altéré conduira à la mort de la cellule, mais si elle est en bonne
|   |   forme elle peut être devenue assez mature pour déclencher sa division **
|   |   Déterminer si la cellule survit ou disparaît par mort ou division
|   |   Mettre à jour l'environnement en fonction de ce qui est consommé et excrété par la
|   |   cellule ou libéré à la mort de la cellule
|   Fin Pour
Fin tant que
```

Algorithme IV.7 : Algorithme de principe du cycle de croissance des cellules.

La division peut être engagée dès que la cellule aura atteint un état de maturité qui le lui autorisera. La maturité d'une cellule est évaluée par un paramètre qui évolue à chaque itération des fonctions cellulaires. Cela représente le travail nécessaire pour constituer tout le matériel cellulaire et les réserves énergétiques destinés aux deux cellules filles.

Bien sûr l'état interne a une influence sur l'efficacité de ce travail et donc sur la façon dont la maturité progresse. La façon dont évolue la maturité est appelée à être affinée pour que son évaluation soit faite par rapport aux excès d'énergie et de matière qui peuvent être dégagés lorsque les quantités consommées le permettent, et donc si les besoins ont été satisfaits. Une cellule qui meurt sera recyclée et ses constituants retourneront dans l'environnement. Pour l'instant nous ne considérons que des nutriments qui sont ainsi relâchés dans le milieu, mais ce recyclage est appelé à être affiné pour gérer différents types de molécules qui seront ainsi

rendues à l'environnement, des nutriments, mais aussi des déchets voire des éléments toxiques.

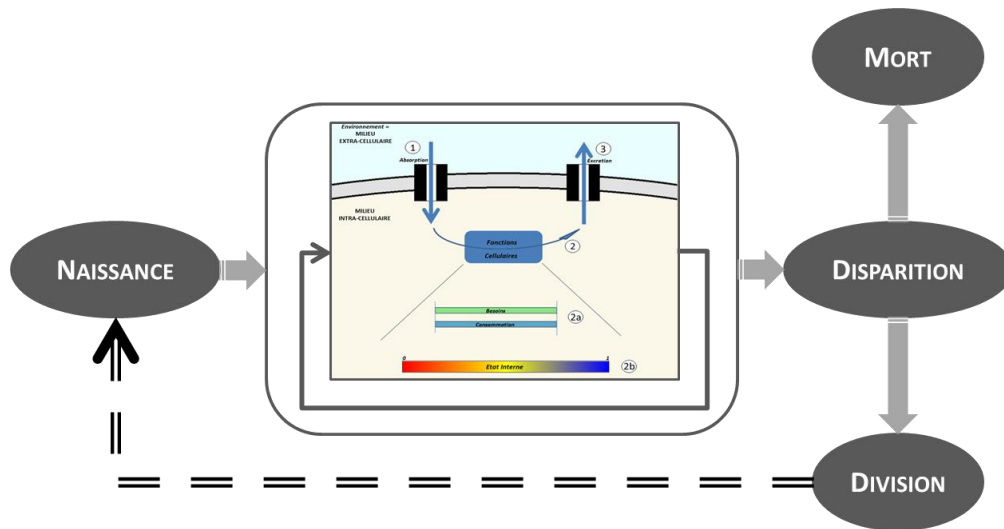


Figure IV.16 : Fonctionnement d'une cellule dans l'automate.

Lors de la naissance d'une cellule, ses caractéristiques sont initialisées. En vertu des lois de la génétique, ses propriétés sont définies en fonction de l'espèce à laquelle elle appartient, modulées par les caractéristiques de la cellule mère et sur lesquelles nous allons rajouter de la variabilité individuelle.

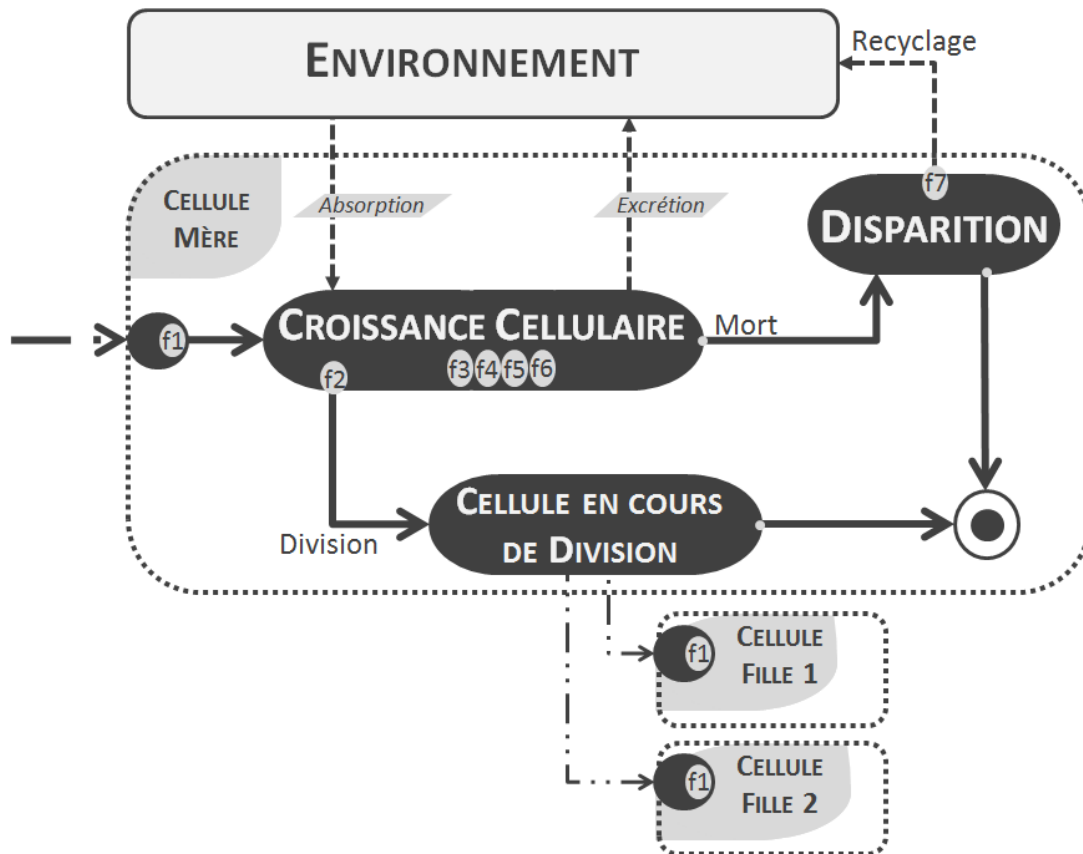


Figure IV.17 : Schéma du principe du fonctionnement du modèle.

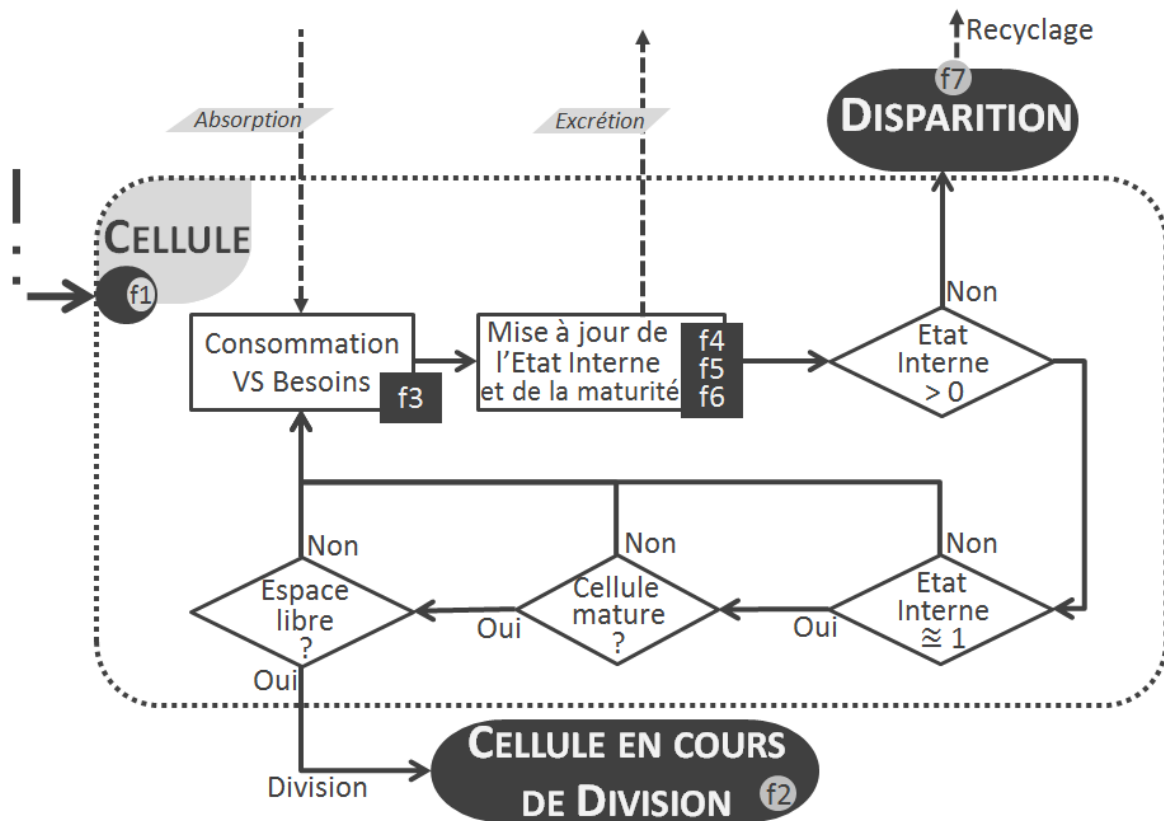


Figure IV.18 : Fonctionnement du modèle de cellule.

Avec tout ceci, le principe du fonctionnement du modèle peut se schématiser comme montré à la Figure IV.17. Et plus précisément, le fonctionnement des cellules peut se schématiser comme montré à la Figure IV.18.

IV.2.2. PARAMETRES DU MODELE

À partir de ces hypothèses de modélisation nous avons construit un modèle de cellule. Pour cela nous avons fait le choix d'un paramétrage en trois niveaux.

Le premier niveau de paramètres permet de rattacher une cellule à une espèce, ces paramètres vont représenter les valeurs moyennes rencontrées chez une espèce, ce seront les paramètres spécifiques, le second niveau permettra de caractériser les individus séparément, ce niveau nous permet ainsi d'introduire la variabilité interindividuelle, ce seront les paramètres individuels et enfin le troisième niveau de paramètres correspondra à des valeurs courantes, c'est-à-dire ce qui sera réellement à chaque instant en fonction de l'influence de l'environnement.

Ceci se traduit dans le modèle par la définition de douze paramètres, six paramètres spécifiques et six paramètres individuels, et de six variables. Chaque cellule est instanciée avec une copie de chacun.

Les caractéristiques de l'espèce se définissent à travers six paramètres spécifiques. En caractéristiques de l'espèce ils sont bien sûr valables et instanciés chez chaque cellule

appartenant à l'espèce considérée. Ils correspondent à moyennes auxquelles se réfèrent tous les individus d'une espèce, par conséquent, une variance est associée à chacun. Ce sont :

- $\text{MAX_CONSUMPTION} \in \mathbb{R}^+ =$ Valeur moyenne de la capacité de consommation d'une cellule ;
- $\text{REQUIREMENTS} \in \mathbb{R}^+ =$ Valeur moyenne des besoins par pas de simulation ;
- $\text{MIN_CYCLE_LENGTH} \in \mathbb{R}^+ =$ Valeur moyenne des durées minimales du cycle cellulaire en conditions optimales ;
- $\text{MAX_UPTAKE} \in \mathbb{R}^+ =$ Valeur moyenne de la capacité maximale d'extraction de ressources de la cellule ;
- $\text{MAX_OUTTAKE} \in \mathbb{R}^+ =$ Valeur moyenne de la capacité maximale d'excrétion de la cellule ;
- $\text{RELEASED_COMPOUNDS} \in \mathbb{R}^+ =$ Valeur moyenne de la quantité d'éléments libérés dans l'environnement lors de la mort d'une cellule. Ces éléments peuvent être nutritifs ou toxiques.

La variance pour chacun de ces paramètres est notée respectivement $\sigma^2_{\text{MAX_CONSUMPTION}}$; $\sigma^2_{\text{REQUIREMENTS}}$; $\sigma^2_{\text{MIN_CYCLE_LENGTH}}$; $\sigma^2_{\text{MAX_UPTAKE}}$; $\sigma^2_{\text{MAX_OUTTAKE}}$; $\sigma^2_{\text{RELEASED_COMPOUNDS}}$. C'est pour l'instant une distribution normale qui a été appliquée sur chacun de ces paramètres, mais le choix de la distribution statistique doit être étudié plus spécifiquement.

Les six paramètres individuels sont évalués en fonction des paramètres qui caractérisent l'espèce. Ils sont évalués pour chaque individu au moment de sa naissance permettant de gérer la notion de variabilité interindividuelle :

- $\text{max_consumption}_i \sim \mathcal{N}(\text{MAX_CONSUMPTION}, \sigma^2_{\text{MAX_CONSUMPTION}}) =$ Capacité maximale de consommation d'une cellule i ;
- $\text{requirements}_i \sim \mathcal{N}(\text{REQUIREMENTS}, \sigma^2_{\text{REQUIREMENTS}}) =$ Besoins d'une cellule i par pas de simulation ;
- $\text{min_cycle_length}_i \sim \mathcal{N}(\text{MIN_CYCLE_LENGTH}, \sigma^2_{\text{MIN_CYCLE_LENGTH}}) =$ Longueur du cycle de la cellule i en conditions optimales ;
- $\text{max_uptake}_i \sim \mathcal{N}(\text{MAX_UPTAKE}, \sigma^2_{\text{MAX_UPTAKE}}) =$ Capacité maximale d'extraction de ressources de la cellule i ;
- $\text{max_outtake}_i \sim \mathcal{N}(\text{MAX_OUTTAKE}, \sigma^2_{\text{MAX_OUTTAKE}}) =$ Capacité maximale d'excrétion de la cellule i ;
- $\text{released_coumpounds}_i \sim \mathcal{N}(\text{RELEASED_COMPOUNDS}, \sigma^2_{\text{RELEASED_COMPOUNDS}}) =$ Quantité d'éléments libérés dans l'environnement lors de la mort de la cellule i .

Pour finir, les six variables sont évaluées à chaque instant à partir des paramètres propres à l'individu pour décrire les valeurs courantes réellement observables à tout moment au sein de la cellule en fonction de son état et de son environnement. Les 5 premières sont des variables réelles continues, la dernière est un booléen.

- $\text{consumption}_i \in [0, \text{max_consumption}_i] =$ Capacité courante de consommation de la cellule i ;
- $\text{cycle}_i \in \mathbb{R}^+ =$ État courant de l'avancement du cycle cellulaire pour la cellule i ;
- $\text{uptake}_i \in [0, \text{max_uptake}_i] =$ Capacité courante d'extraction de la cellule i ;
- $\text{outtake}_i \in [0, \text{max_outtake}_i] =$ Capacité courante d'excrétion de la cellule i ;

- $Ei_i \in [0, 1]$ = État interne courant de la cellule i ;
- $readiness_i \in \{true ; false\}$ = État de maturité de la cellule i , cette valeur vaut « true » si la cellule est prête à se diviser ;

Les calculs sur les paramètres et variables des cellules lors des itérations de l'absorption, de l'extraction, des fonctions cellulaires font appel à des fonctions, f1 à f7 dont la définition est à faire et à affiner au cas par cas.

- **f1** : loi de probabilité qui relie les paramètres moyens connus pour la population (espèce) aux paramètres de l'individu (cellule), elle est utilisée lors de la création d'un individu. La même fonction est appliquée à chacun des paramètres, mais là aussi, affiner le modèle passera par une gestion plus fine de la variabilité interindividuelle selon les caractéristiques étudiées. Elle reçoit en argument un caractère spécifique et sa variance et donne la valeur du caractère individuel.
- **f2** : relation entre état interne, environnement et paramètres de la cellule pour évaluer ses capacités à extraire des ressources de son environnement et à les consommer au pas de temps donné pour une cellule en cours de division. Cette fonction dira si la division s'est correctement déroulée, et si c'est le cas, elle déclenchera l'initialisation des 2 cellules filles. Elle rend une valeur booléenne pour indiquer si la division a pu être menée à terme et les valeurs courantes d'uptake et de consommation pour une cellule en cours de division. Pour définir ces valeurs, elle prend en arguments les valeurs de Ei et des paramètres individuels $max_consumption$ et max_uptake . Elle est modulée par les propriétés de l'environnement,
- **f3** : relation entre état interne, environnement et paramètres individuels de la cellule pour évaluer ses capacités à extraire des ressources de son environnement et à les consommer au pas de temps donné. Cette fonction s'apparente à f2 si ce n'est que f2 est spécifique du processus de division. La fonction f3 est utilisée pour chaque opération de la vie de la cellule. Ainsi, elle rend comme résultat une valeur booléenne qui indique si la cellule a survécu à ce pas de temps et les valeurs courantes d'uptake et de consommation. Pour ce faire, elle considère en argument les valeurs de Ei ainsi que les paramètres individuels $max_consumption$ et max_uptake . Le résultat de cette fonction est également modulé par les propriétés de l'environnement (voir paragraphe IV.4.1, page 155),
- **f4** : capacité de la cellule à restaurer son état interne en fonction de sa consommation, de son état interne. Cette fonction a pour résultat la valeur de Ei et « ce qui reste » une fois les besoins cellulaires assouvis et l'état interne restauré. Ce calcul est réalisé en prenant en argument la valeur de Ei , de requirements et de consommation. Cette fonction est elle aussi assujettie aux propriétés de l'environnement.
- **f5** : relation entre la dégradation de l'état interne et la carence subie. Cette fonction est appelée en cas de carence dans la consommation cellulaire. Elle permet d'estimer l'importance de la carence et d'en répercuter l'effet sur la valeur de Ei . Cette fonction prend en argument les valeurs de consommation, de Ei et de requirements. Son résultat est modulé par les propriétés de l'environnement.
- **f6** : relation entre l'état interne de la cellule et l'avancement de son cycle cellulaire. Cette fonction permet d'évaluer la valeur de cycle en prenant en

argument la valeur de E_i . Les propriétés de l'environnement a aussi une influence sur cette fonction.

- **f7** : Relation entre la modification de l'environnement et les constituants cellulaires lorsqu'une cellule meurt. La quantité d'éléments ainsi rendus à l'environnement est calculée en fonction de `released_coumpounds` et de E_i .

Nous avons alors pour décrire le modèle de cellule l'algorithme suivant :

```

Pour chaque cellule  $i$  faire
/   ** Initialisation des variables **
|   Booléen survie  $\leftarrow$  true
|   divisionProcess  $\leftarrow$  false
|   Réel excrétion  $\leftarrow$  0
|   Réel surplus  $\leftarrow$  0
|   Réel excédent  $\leftarrow$  0
|   Réel manque  $\leftarrow$  0
/   ** Évaluation des paramètres courants de la cellule **
|   Si readiness $_i$  = true alors
|       |   (divisionProcess, uptake $_i$ , consumption $_i$ )  $\leftarrow$  f2 ( $E_i$ , environnement,
|       |   max_consumption $_i$ , max_uptake $_i$ )
|       |       Si divisionProcess = true alors
|       |           |   initialisation de la cellule  $i$ 
|       |           |   initialisation de la cellule fille
|       |       Fin si
|       Sinon
|           |   (uptake $_i$ , consumption $_i$ )  $\leftarrow$  f3 ( $E_i$ , environnement, max_consumption $_i$ , max_uptake $_i$ )
|       Fin si
/   ** On compare la capacité d'absorption évaluée de la cellule à la quantité disponible dans l'environnement **
|       Si uptake $_i$  > environnement $_i$  alors
|           |   uptake $_i$   $\leftarrow$  environnement $_i$ 
|       Fin si
/   ** On compare la capacité d'absorption évaluée à la capacité maximale de consommation de la cellule **
|       Si uptake $_i$  > max_uptake $_i$  alors
|           |   uptake $_i$   $\leftarrow$  max_uptake $_i$ 
|       Fin si
/   ** On compare la quantité absorbée à ce qui peut être consommé par la cellule **
|       Si uptake $_i$  < consumption $_i$  alors
|           |   consumption $_i$   $\leftarrow$  uptake $_i$ 
|           |   excrétion  $\leftarrow$  0
|       Sinon
|           |   excrétion  $\leftarrow$  excrétion + (uptake $_i$  - consumption $_i$ )
|       Fin si
/   ** L'état interne et la position dans le cycle cellulaire sont mis à jour en fonction de ce qui a été consommé **
|       Si consumption $_i$  > requirements $_i$  alors
|           |       Si consumption $_i$  > max_consumption $_i$  alors
|           |           |   surplus  $\leftarrow$  surplus + (consumption $_i$  - max_consumption $_i$ )
|           |           |   consumption $_i$   $\leftarrow$  max_consumption $_i$ 
|           |       Fin si

```

```

|      |      excédent  $\leftarrow$  consumptioni – (requirementsi)
|      |      Si Eli < 1 alors
|      |      |      (Eli, excédent)  $\leftarrow$  f4 (Eli, excédent, environnement)
|      |      Fin si
|      |      excrétion  $\leftarrow$  excrétion + excédent + surplus
|      |      Sinon
|      |      |      manque  $\leftarrow$  requirementsi – consumptioni
|      |      |      Eli  $\leftarrow$  f5 (Eli, manque, environnement)
|      |      |      readinessi  $\leftarrow$  false
|      |      Fin si
|      |      outtakei  $\leftarrow$  excrétion
|      |      Si cyclei > min_cycle_lengthi et Eli  $\geq$  (1- $\epsilon$ ) alors
|      |      |      readinessi  $\leftarrow$  true
|      |      Fin si
|      |      Si Eli  $\leq$  ( $\epsilon$ ) alors
|      |      |      survie  $\leftarrow$  false
|      |      |      Eli  $\leftarrow$  0
|      |      |      environnement  $\leftarrow$  f7 (environnement, released_compoundsi, Eli)
|      |      Fin si
|      |      cyclei  $\leftarrow$  f6 (Eli, environnement)
|      Fin pour

```

Algorithme IV.8 : Algorithme détaillé du modèle.

IV.2.3. DEVENIR DES CELLULES FILLES.

Nous considérons l'hypothèse que la division n'est pas orientée. Mais c'est une possibilité qui reste ouverte et à envisager pour certains types de bactéries (genre Strepto- qui forment des chaînettes).

Avec cette hypothèse, une cellule qui se divise a autant de chance de « grandir » dans toutes les directions de l'espace, du moins tant que tout l'espace autour est libre. Sinon, les espaces libres « acquièrent » plus de poids et orientent ou limitent les directions possibles pour la division. De plus, dans l'hypothèse de simulation d'une culture en milieu liquide (en termes d'effectif), soit un milieu agité, la direction de la division n'a pas vraiment de signification.

Dans le modèle cela se traduit par le fait qu'une cellule qui est prête à se diviser consulte son voisinage et « choisit aléatoirement et de manière équiprobable » une maille vide qui devient cellule fille. La maille qui contenait la cellule mère devient l'autre cellule fille. La taille du voisinage décrit la capacité de propagation des cellules lors de la division.

IV.2.4. SUPERINDIVIDU

Ce modèle a été implémenté sur la plateforme sous la forme d'un automate cellulaire avec pour contrainte 1 cellule par maille. Ceci s'est traduit par une trop forte contrainte sur l'espace qui limitait l'effectif maximum qui pouvait être atteint et la cinétique de la croissance, ce sera détaillé dans le chapitre Résultat du présent document (paragraphe V.4.2, page 175). Pour relâcher cette contrainte nous avons fait appel à la notion de

« superindividu » (Scheffer et *al.*, 1995 [135]), c'est-à-dire qu'un individu ne représente plus une cellule unique, mais un groupe, une sous-population. Pour ce faire une septième variable a été rajoutée :

- cellulesParMaille_i ∈ ℕ = nombre d'individus représentés dans une même maille

Lors de la division cellulaire, nous vérifions si une la cellule mère disposait d'un espace adjacent libre pour accueillir la cellule fille, cette hypothèse dérivait de l'inhibition de contact qui existe chez les cellules des organismes animaux. Cette limitation n'a pas lieu d'être pour la modélisation de bactéries. Cette notion de superindividu nous permet de conserver cette hypothèse pour garder une certaine généricité au modèle en fixant par exemple une valeur maximale au nombre de cellules représentées par un superindividu, tout en le rendant plus cohérent avec la modélisation de microorganismes ; il permet de considérer qu'une cellule qui se divise dispose toujours d'espace disponible à proximité pour accueillir la cellule fille, si cet espace n'est pas dans une maille adjacente, il est dans la maille qui accueille la cellule mère. Cette dernière maille reçoit alors les deux cellules filles.

Avec cette variable, la cellule est maintenant un « superindividu » qui représente de 1 à n individus. Les fonctions f2, f3, et f7 dépendent alors en plus de ce nouveau paramètre. L'Algorithme IV.8 devient alors l'Algorithme IV.9.

L'algorithme devient alors :

```

Pour chaque cellule i faire
/   ** Initialisation des variables **
|   Booléen survie ← true
|   divisionProcess ← false
|   Réel excrétion ← 0
|   Réel surplus ← 0
|   Réel excédent ← 0
|   Réel manque ← 0
/   ** Évaluation des paramètres courants de la cellule **
|   Si readinessi = true alors
|       |   (divisionProcess, uptakei, consumptioni) ← f2 (Ei, environnement,
|       |   max_consumptioni, max_uptakei, cellulesParMaillei)
|       |   Si divisionProcess = true alors
|       |       |   initialisation de la cellule i
|       |       |   initialisation de la cellule fille
|       |   Fin si
|   Sinon
|       |   (uptakei, consumptioni) ← f3 (Ei, environnement, max_consumptioni, max_uptakei,
|       |   cellulesParMaillei)
|   Fin si
/   ** On compare la capacité d'absorption évaluée de la cellule à la quantité disponible dans
l'environnement **
|   Si uptakei > environnementi alors
|       |   uptakei ← environnementi
|   Fin si
/   ** On compare la capacité d'absorption évaluée à la capacité maximale de consommation
de la cellule **
|   Si uptakei > max_uptakei alors

```



```

|      uptakei ← max_uptakei
|      Fin si
|      ** On compare la quantité absorbée à ce qui peut être consommé par la cellule **
|      Si uptakei < consommationi alors
|      |      consommationi ← uptakei
|      |      excréation ← 0
|      Sinon
|      |      excréation ← excréation + (uptakei – consommationi)
|      Fin si
|      ** L'état interne et la position dans le cycle cellulaire sont mis à jour en fonction de ce qui
|      a été consommé **
|      Si consommationi > (requirementsi * cellulesParMaillei) alors
|      |      Si consommationi > max_consommationi alors
|      |      |      surplus ← surplus + (consommationi – max_consommationi)
|      |      |      consommationi ← max_consommationi
|      |      Fin si
|      |      excédent ← consommationi – (requirementsi * cellulesParMaillei)
|      |      Si Eli < 1 alors
|      |      |      (Eli, excédent) ← f4 (Eli, excédent, environnement)
|      |      Fin si
|      |      excréation ← excréation + excédent + surplus
|      Sinon
|      |      manque ← requirementsi – (consommationi/cellulesParMaillei)
|      |      Eli ← f5 (Eli, manque, environnement)
|      |      readinessi ← false
|      Fin si
|      outtakei ← excréation
|      Si cyclei > min_cycle_lengthi et Eli ≥ (1-ε) alors
|      |      readinessi ← true
|      Fin si
|      Si Eli ≤ (ε) alors
|      |      survie ← false
|      |      Eli ← 0
|      |      environnement ← f7 (environnement, released_compoundsi, Eli, cellulesParMaillei)
|      Fin si
|      cyclei ← f6 (Eli, environnement)
Fin pour

```

Algorithme IV.9 : Algorithme détaillé du modèle modifié pour intégrer la notion de superindividu

Les fonctions f2, f3, et f7 dépendent alors en plus de ce nouveau paramètre.

IV.3. DEFINITION DES VALEURS DES PARAMETRES

Les comportements du modèle et de la plateforme ont été validés et vérifiés avec des jeux de paramètres simples et arbitraires, à partir de configurations initiales définies, sur différentes configurations matérielles et logicielles.

Les paramètres du modèle ont ensuite été affinés nous verrons comment.

Nous avons cherché des valeurs chiffrées du comportement biologique des fonctions cellulaires en particulier sur la consommation du glucose par *E. coli*. La thèse de Cécile Nicolas (Nicolas, 2007 [136]) a retenu notre attention. Elle décrit les quantités consommées de glucose par gramme de bactéries et par heure, ainsi que les conditions d'expérimentations qui ont donné ces résultats.

En culture en milieu minimum synthétique, avec 2 g/l de glucose, ils obtiennent une consommation de glucose de 7,15 mmol/g/h. Ce qui signifie que dans ces conditions expérimentales, en 1 h de temps, 1g de bactéries ont consommé 7,15 mmol soit $7,15 \times 10^{-3}$ mol de glucose.

Ils ne s'intéressent pas au nombre de bactéries mais à la biomasse, et le facteur qu'ils utilisent pour corréliser biomasse et DO lors des suivis de croissance dit que 1 unité de DO correspond à 0,38 g de biomasse bactérienne sèche par litre.

Selon les estimations, 1 de DO correspondent de 10^7 à 10^9 cellules par millilitre.

Pour la suite des calculs nous ne conserverons que l'hypothèse 10^9 cellules par millilitre pour 1 de DO, Nous justifierons ce choix dans le paragraphe IV.3.3 (page 154).

IV.3.1. CONSOMMATION PAR BACTERIE

La consommation en glucose a été relevée à $7,15 \times 10^{-3}$ mol/g/h.

Nous venons de déterminer qu'1 g de bactérie correspond à $2,63 \times 10^{12}$ bactéries. On peut donc déduire qu'une bactérie consomme par heure $2,718 \times 10^{-15}$ mol de glucose.

En vertu du nombre d'Avogadro, on peut traduire cette consommation en la consommation en 1 h par une bactérie de $1,636 \times 10^9$ molécules de glucose.

Une autre estimation qui conforte ce calcul est donnée par Kondev et Theriot. En se basant sur le nombre d'atomes de carbones, ils ont estimé que pour qu'une cellule d'*E. coli* double sa taille et se prépare à se diviser, il lui faudra consommer 2×10^9 molécules de glucose (Philips, Kondev et Theriot, 2008 [137], [138]).

IV.3.2. ENVIRONNEMENT MODELISE

Les simulations que nous avons réalisées plaçaient les bactéries modélisées dans une matrice carrée de taille 500x500. Sans utiliser la notion de superindividu, l'hypothèse était faite que chaque maille de cette matrice ne pouvait héberger qu'une seule cellule.

En supposant qu'une cellule remplit complètement une maille, on peut estimer le volume représenté par une maille égal au volume d'une cellule.

Une matrice de 500x500 comprend 250 000 mailles. Si une maille correspond précisément au volume d'une cellule, la matrice modélisée représente exactement le volume de 250 000 cellules.

E. coli est un bacille, c'est-à-dire une bactérie qui a une forme de bâtonnet dont les bouts sont arrondis. Sa longueur est de 1 à 2 μm , et son diamètre est de 0,5 à 1 μm . Le volume moyen d'une cellule d'*E. coli* est ainsi estimé de 0,6 à 1,5 μm^3 .

Avec l'hypothèse la plus basse, 0,6 μm^3 par cellule, on obtient une matrice qui représente 150 000 μm^3 , soit 150 000 fL, soit 150 pL ; avec l'hypothèse haute, 1,5 μm^3 par cellule, on a une matrice qui représente 375 000 μm^3 soit 375 000 fL ou encore, 375 pL.

Si l'on considère que les conditions expérimentales utilisaient un milieu à une concentration en glucose de 2 g/L. Le glucose est un sucre dont la formule chimique est $\text{C}_6\text{H}_{12}\text{O}_6$, qui possède une masse molaire de 180 g/mol (C = 12 g/mol, H = 1 g/mol ; O = 16 g/mol). Une concentration de 2 g/L correspond ainsi à 11×10^{-3} mol/L, c'est-à-dire 11×10^{-14} mol/pL.

En vertu du nombre d'Avogadro, on a donc 6.69×10^9 molécules de glucose par picolitre de milieu. Ainsi, selon le volume de la cellule considéré nous devons introduire dans le milieu modélisé $6,688 \times 10^9$ molécules de glucose par picolitre

- soit $150 \times 6,688 \times 10^9 = 1,003 \times 10^{12}$ molécules dans l'hypothèse où une cellule représente 0,6 μm^3 ;
- soit $375 \times 6,688 \times 10^9 = 2.508 \times 10^{12}$ molécules dans l'hypothèse où une cellule représente 1,5 μm^3 .

Ce qui correspondra respectivement à $4,01 \times 10^6$ et $1,00 \times 10^7$ molécules par mailles.

IV.3.3. DENSITE DE POPULATION BACTERIENNE DONNANT UNE DO DE 1

On estime que la masse de protéines contenues dans une cellule de *E. coli* est de 0,15 pg (Bionumbers ; [137]), et on estime que les protéines représentent 50% de la masse sèche d'une bactérie (Bionumbers ; [137]). Ceci nous donne que la masse sèche peut être estimée à d'une cellule de *E. coli* est d'environ 0,3 pg.

Dans ces travaux, C. Nicolas a mesuré que lorsque la DO valait 1, cela correspondait à 0,38 g de biomasse sèche par litre de culture. De là on peut calculer le nombre de bactéries que représentent 0,38 g de matière sèche :

$$nb_{bactéries} = \frac{0,38}{0,3 * 10^{-12}} \approx 1,27 * 10^{12}$$

Avec 0,38 g de biomasse sèche par litre de culture, cela correspond à $1,27 \times 10^{12}$ bactéries, soit $1,27 \times 10^9$ cellules par millilitre.

On estime couramment que 1 de DO représente de 10^7 à 10^9 bactéries par millilitre, avec ce que nous venons de calculer, nous ne conserverons que l'estimation à 10^9 cellules/mL pour définir l'ordre de grandeur de la population que nous devons atteindre au cours d'une simulation avec ces données expérimentales.

IV.4. FONCTIONNEMENT ET IMPLEMENTATION

Pour le fonctionnement du modèle au sein de l'automate cellulaire géré par la plateforme nous avons introduit des paramètres et variables supplémentaires. Insistons sur le fait qu'ils ne sont là que pour la mise en œuvre de l'implémentation et ne changent pas le modèle.

IV.4.1. PARAMETRES DE L'AUTOMATE CELLULAIRE

Une Maille présente cinq données :

- $Status_i \in \{0 ; 1\}$ = État courant de la maille i , il désigne si la maille i est vivante ou morte ;
- $New_Status_i \in \{0 ; 1\}$ = État mis à jour de la maille i , il est calculé en fonction de $Status_i$ et des Status contenus dans N_i ;
- $Updated_i \in \{0 ; 1\}$ = Ce paramètre vaut 1 si la maille i a été mise à jour ;
- N_i = Liste des mailles dans le voisinage de la maille i , et de leur état. Il dépend de la topologie de l'automate et du système de voisinage choisi ;
- $Environnement_i$ = Liste des caractéristiques de l'environnement présent dans la maille i . Elle contient les propriétés physiques (température, pH, ...) et les propriétés chimiques (liste des constituants et leur concentration respective) de l'environnement. La taille de cette liste dépend du système de voisinage choisi ;

Si le Status d'une maille vaut 1, cela signifie qu'une cellule vivante occupe cette maille. Cette cellule possède les paramètres et variables listés précédemment : Il en est de même pour le New_Status .

Chaque cellule est en plus pourvue de 2 variables qui seront utilisées dans les calculs sur la population.

- $birthDate_i \in \mathbb{N}$ = Date de naissance de la cellule i ;
- $disDate_i \in \mathbb{N}$ = Date de disparition de la cellule i , soit parce qu'elle meurt, soit parce qu'elle se divise et donne naissance à 2 nouvelles cellules.

IV.4.2. GESTION ET SUIVI DE LA SIMULATION

Pour suivre les événements pas à pas, maille par maille nous avons mis en place des variables gérées par la plateforme qui écoutent le comportement du modèle.

Ainsi, nous avons :

- Des variables de calcul pour chaque maille :
 - o $nbEmptyCells_i \in \mathbb{N}$ = Nombre de mailles libres dans le voisinage (N_i) de la cellule i , il est compris entre 0 si aucune maille du voisinage de la cellule i n'est vide et le cardinal de N_i si toutes les mailles du voisinage de la cellule i sont libres ;
 - o $emptyCells_i$ = Liste de cellules qui stocke les mailles libres présentes dans le voisinage de la cellule i . Cette liste aura une taille donnée par $nbReadyCells_i$.

- $MaI_i \in \{\text{true} ; \text{false}\}$ = Dans le cas où la cellule i s'est divisée, cette valeur vaut « true » et désigne le fait qu'une maille vide de la liste emptyCells_i accueille une nouvelle cellule qui doit être initialisée en fonction des paramètres de la cellule i ;
- daughterCell_i = Maille désignée pour accueillir la nouvelle cellule issue de la division de la cellule i . Elle est désignée aléatoirement dans la liste emptyCells_i reçoit une nouvelle cellule initialisée en fonction des paramètres de la cellule i ;
- $\text{releasedCompounds}_i \in [0, \text{RELEASED_COMPOUNDS}]$ = Quand une cellule meurt, tout ou partie de son matériel est rendu à l'environnement en fonction de l'environnement et de l'état de la cellule i ;
- $\text{survie}_i \in \{\text{true} ; \text{false}\}$ = Cette variable prend la valeur « true » si la fonction cellulaire a abouti à la survie de la cellule i ;
- $\text{divisionProcess}_i \in \{\text{true} ; \text{false}\}$ = Cette variable prend la valeur « true » si la cellule i est entrée en division et que cette fonction a généré 2 nouvelles cellules correctement initialisées.
- Des variables de calcul globales qui permettent le suivi de la simulation :
 - $\text{timeStep} \in \mathbb{N}$ = Cette valeur compte les itérations de la simulation. Elle sera utilisée pour définir la date de naissance (birthDate_i), et la date de disparition (disDate_i) d'une cellule i .
 - $\text{nbCellulesDivisees} \in \mathbb{N}$ = Cette variable stocke tout au long de la simulation le nombre de cellules qui entrent dans le processus de division, de 0 au début, elle est incrémentée de 1 à chaque cellule qui entre dans ce processus ;
 - $\text{totalCellDivisionAge} \in \mathbb{R}^+$ = Dans cette variable, l'âge total des cellules qui entrent en division est évalué. Initialisée à 0, elle est ensuite incrémentée de la valeur cycle de chaque cellule qui entre dans le processus de division. À la fin de la simulation, diviser cette variable par $\text{nbCellulesDivisees}$ donnera l'âge moyen des cellules qui entrent dans le processus de division ;
 - $\text{totalCellLifeLength} \in \mathbb{R}^+$ = Cette variable est initialisée à 0 et est incrémentée lorsqu'une cellule disparaît du temps qu'elle aura existé. C'est-à-dire que lorsqu'une cellule i disparaît parce qu'elle entre dans un processus de division, on calcule la différence entre disDate_i et birthDate_i et cette valeur est ajoutée à $\text{totalCellLifeLength}$. À la fin de la simulation, en divisant cette variable par $\text{nbCellulesDivisees}$ on obtiendra le temps de présence moyen d'une cellule entre sa naissance et son entrée dans le processus de division.

Rappelons que pour modéliser différentes espèces, il faut modifier les paramètres MAX_CONSUMPTION , REQUIREMENTS , MIN_CYCLE_LENGTH , MAX_UPTAKE , MAX_OUTTAKE , $\text{RELEASED_COMPOUNDS}$, et que changer le milieu ou les conditions de culture se fera en modifiant les valeurs des éléments de Environnement.

IV.4.3. INITIALISATION D'UNE CELLULE

Nous avons choisi de procéder de deux façons distinctes pour initialiser les cellules, la première n'est employée que lors de la création de l'automate, la seconde est employée pour la naissance de chaque nouvelle cellule au cours de la simulation. Les cellules qui naissent au cours de la simulation ont en effet une cellule mère identifiée sur laquelle se baser pour calculer leurs paramètres, ce qui n'est pas possible pour la première génération. Cette fonction spécifique d'initialisation des cellules lors de la création de la population initiale a pour but de mimer l'existence d'une population antérieure.

**IV.4.3.1. INITIALISATION D'UNE CELLULE DE LA
GENERATION 0**

```

max_consumptioni ← f1 (MAX_CONSUMPTION)
requirementsi ← f1 (REQUIREMENTS)
min_cycle_lengthi ← f1 (MIN_CYCLE_LENGTH)
max_uptakei ← f1 (MAX_UPTAKE)
max_outtakei ← f1 (MAX_OUTTAKE)
consumptioni ∈ [0 ; max_consumptioni] : capacité courante de consommation de la cellule i
stocki ← 0
cyclei ← 0
uptakei ← 0
outtakei ← 0
EIi ← 1
readinessi ← false
releasedCompoundsi ← 0,25
birthDatei ← 0
disDatei ← 0

```

Algorithme IV.10 : Algorithme de la fonction d'initialisation d'une cellule de la génération 0.

IV.4.3.2. INITIALISATION A LA NAISSANCE D'UNE CELLULE

```

max_consumptioni ← f1(max_consumptionmotherCell)
requirementsi ← f1(requirementsmotherCell)
min_cycle_lengthi ← f1(min_cycle_lengthmotherCell)
max_uptakei ← f1(max_uptakemotherCell)
max_outtakei ← f1(max_outtakemotherCell)
consumptioni ← 0
cyclei = ← 0
uptakei ← 0
outtakei ← 0
EIi ← 1
readinessi ← false
releasedCompoundsi ← 0,25
birthDatei ← timeStep + 1
disDatei ← 0

```

Algorithme IV.11 : Algorithme d'initialisation d'une cellule.

IV.4.4. FONCTIONS CELLULAIRES

Nous avons défini une fonction qui régit les échanges entre la cellule et son environnement et aboutit à la mise à jour de ses paramètres. Le résultat de cette fonction indique à l'automate si la cellule a survécu ou non.

Ici aussi nous avons pris le parti de dupliquer l'implémentation de ce processus. La première version consiste en la version « normale », celle utilisée par chaque cellule à chaque itération. La seconde version correspond au cas particulier d'une cellule en cours de division. En effet, il nous est apparu important de pouvoir manipuler plus finement cet état transitoire qui conduit d'une cellule mère à deux cellules filles.

Remarque : nous faisons l'hypothèse que le paramètre environmentFactor n'est pris en compte qu'au moment du calcul de la valeur courante des variables de la cellule et qu'il n'a pas d'influence sur les paramètres de la cellule.

IV.4.4.1. FONCTION REGISSANT LE DEROULEMENT DU CYCLE CELLULAIRE

*** Initialisation des variables ***

Booléen survie \leftarrow true

Réel excrétion \leftarrow 0

Réel surplus \leftarrow 0

Réel excédent \leftarrow 0

Réel manque \leftarrow 0

Réel environmentFactor \leftarrow 1

Réel degradationRate \leftarrow 0,15

$\varepsilon \leftarrow 10^{-6}$

outtake_i \leftarrow 0

uptake_i \leftarrow max_uptake_i * environmentFactor * E_i * 1

consumption_i \leftarrow E_i * max_consumption_i * environmentFactor * 1

*** Absorption des nutriments & Maintenance ***

*** On compare l'absorption évaluée à la quantité disponible dans l'environnement ***

Si uptake_i > environnement_i **alors**

| uptake_i = environnement_i

Fin si

*** On compare la capacité d'absorption évaluée à la capacité maximale de consommation ***

Si uptake_i > max_uptake_i **alors**

| uptake_i \leftarrow max_uptake_i

Fin si

*** On compare la quantité absorbée à ce qui peut être consommé par la cellule ***

Si uptake_i < consumption_i **alors**

| consumption_i \leftarrow uptake_i

Sinon

| excrétion \leftarrow excrétion + (uptake_i – consumption_i)

Fin si

*** L'état interne et la position dans le cycle cellulaire sont mis à jour en fonction de ce qui a été consommé par rapport aux besoins de la cellule ***

```

Si consumptioni > requirementsi alors
|   Si consumptioni > max_consumptioni alors
|   |   surplus ← surplus + (consumptioni – max_consumptioni)
|   |   consumptioni ← max_consumptioni
|   Fin si
|   excédent ← excédent + (consumptioni – requirementsi)
|   Si Eli < 1 alors
|   |   Si (excédent*1) < (1-Eli) alors
|   |   |   Eli ← Eli + excédent*1
|   |   |   excédent ← 0
|   |   Sinon
|   |   |   excédent ← (Eli + excédent*1) -1
|   |   |   Eli ← 1
|   |   Fin si
|   Fin si
|   excrétion ← excrétion + surplus + excédent
Sinon
|   manque ← requirementsi – consumptioni
|   Eli ← Eli – manque * degradationRate
|   readinessi ← false
Fin si
outtakei ← excrétion
Si cyclei > min_cycle_lengthi et Eli ≥ (1-ε) alors
|   readinessi ← true
Fin si
Si Eli ≤ 0 alors
|   survie ← false
|   Eli ← 0
Fin si
Cyclei ← cyclei + Eli*1
    
```

Algorithme IV.12 : Algorithme de la fonction gouvernant le cycle cellulaire.

IV.4.4.2. FONCTION REGISSANT LES ECHANGES ENTRE LA CELLULE ET SON ENVIRONNEMENT AU COURS DU PROCESSUS DE DIVISION

Le résultat final indique à l'automate si la division s'est bien passée et dans l'affirmative, qu'il doit maintenant gérer 2 nouvelles cellules filles, l'une ayant pris la place de sa cellule mère l'autre ayant pris place dans une maille qui était libre dans le voisinage N_i de la cellule mère i .

Nous utilisons ici un nouveau paramètre, **growthFactor**, pour obtenir la possibilité de moduler les échanges entre la cellule et son environnement en fonction de la façon dont est considérée la cellule en cours de division. On s'offre ainsi la possibilité de considérer qu'une cellule en cours de division compte, en termes d'échanges avec son environnement, non plus comme une seule cellule, mais comme deux cellules par exemple, avec toutes les variations

intermédiaires, ou au contraire qu'une cellule qui est en cours de division réduit ces échanges avec son environnement pour se concentrer sur sa division.

En nous basant sur ce qui est observé chez les bactéries, nous avons pris le parti de ne pas considérer la survenue d'une carence pendant le processus de division. Ce processus engagé sera mené à terme, les deux cellules filles générées ne percevront le manque qu'au cours de leur vie autonome.

Notons également que cette fonction qui représente le processus de division de la cellule mère en deux cellules filles se déroule sur un pas de temps de la simulation, mais il peut être envisagé de faire durer ce processus sur une durée plus importante.

*** Initialisation des variables ***

Booléen divisionProcess \leftarrow false

Réel excretion \leftarrow 0

Réel surplus \leftarrow 0

Réel excédent \leftarrow 0

Réel environmentFactor \leftarrow 1

Réel growthFactor \leftarrow 2

$\varepsilon \leftarrow 10^{-6}$

outtake_i \leftarrow 0

uptake_i \leftarrow max_uptake_i * environmentFactor * El_i * growthFactor * 1

consumption_i \leftarrow El_i * max_consumption_i * environmentFactor * growthFactor * 1

*** Absorption des nutriments & Maintenance ***

*** On compare la capacité d'absorption évaluée de la cellule à la quantité disponible dans l'environnement ***

Si uptake_i > environnement_i **alors**

| uptake_i \leftarrow environnement_i

Fin si

*** On compare la capacité d'absorption évaluée à la capacité maximale de consommation de la cellule ***

Si uptake_i > (max_uptake_i * growthFactor) **alors**

| uptake_i \leftarrow max_uptake_i * growthFactor

Fin si

*** On compare la quantité absorbée à ce qui peut être consommé par la cellule ***

Si uptake_i < consumption_i **alors**

| consumption_i \leftarrow uptake_i

Si non

| excretion \leftarrow excretion + (uptake_i – consumption_i)

Fin si

*** L'état interne et la position dans le cycle cellulaire sont mis à jour en fonction de ce qui a été consommé par rapport aux besoins de la cellule ***

Si consumption_i > (requirements_i * growthFactor) **alors**

| **Si** consumption_i > (max_consumption_i * growthFactor) **alors**

| | surplus \leftarrow surplus + (consumption_i – (max_consumption_i * growthFactor))

| | consumption_i \leftarrow max_consumption_i * growthFactor

| **Fin si**

| excédent \leftarrow excédent + (consumption_i – requirements_i * growthFactor)

| excretion \leftarrow excretion + surplus + excédent

Fin si

Outtake_i \leftarrow excretion

```
Cyclei ← cyclei + Eli*1  
divisionProcess ← true
```

Algorithme IV.13 : Algorithme de la fonction gouvernant le cycle cellulaire dans le cas particulier d'une cellule en cours de division.

IV.5. CONCLUSION

Suite à l'état de l'art nous avons identifié le besoin de concevoir une plateforme. Au cours de ce chapitre nous avons présenté toutes les étapes de validation qui ont jalonné son élaboration.

Les automates cellulaires testés sur cette plateforme ont fonctionné correctement, que ce soit des modèles d'automates cellulaires de la littérature (le Jeu de la Vie et Wireworld) ou un modèle que nous avons implémenté sous forme d'automate cellulaire (DaisyWorld).

Nous avons également conçu un modèle de cellules biologiques pour tenter de répondre à notre problématique. Nous allons valider son comportement et le mettre en œuvre avec des paramètres significatifs en biologie dans le chapitre suivant.

CHAPITRE V. MISE EN ŒUVRE DU MODELE ET DE LA PLATEFORME : RESULTATS

V.1. Matériel et méthodes	165
V.2. Précision sur le fonctionnement de l'automate cellulaire.....	167
V.3. Reproductibilité.....	168
V.4. Validation du comportement du modèle et de la robustesse des résultats.....	169
V.4.1. Jeux d'essais	169
V.4.1.1. Paramètres.....	169
V.4.1.1.1. Paramètres des cellules	169
V.4.1.1.2. Variables des cellules	169
V.4.1.1.3. Paramètres utilisés dans les fonctions	170
V.4.1.1.4. Campagne de tests	170
V.4.1.2. Croissance observée	171
V.4.1.2.1. Croissance depuis un foyer.....	171
V.4.1.2.2. Croissance à partir de plusieurs foyers	173
V.4.2. Bilan des résultats sans exploiter le superindividu	175
V.4.3. Résultats avec utilisation du superindividu	176
V.5. Utilisation des paramètres issus d'études biologiques	178
V.5.1. Implémentation	178
V.5.2. Résultats.....	178
V.5.3. Phase stationnaire.....	180
V.6. Conclusion.....	183

Au cours de ce chapitre nous allons regarder le comportement du modèle, en commençant par présenter les tests réalisés pour vérifier que son comportement est reproductible, et qu'il est bien celui attendu. Nous ferons auparavant une digression pour apporter une précision sur le fonctionnement de l'automate qu'il nous a fallu prendre en considération. Toutes ces étapes sont réalisées en utilisant un jeu de paramètres simples fixés arbitrairement, avec une variabilité interindividuelle nulle, avec la limite d'une seule cellule par maille (pas de superindividu). Le détail de ces paramètres sera présenté au paragraphe V.4.1.1, page 169 de la présente section.

Nous passerons ensuite à la présentation des résultats obtenus en appliquant des paramètres « biologiques » au modèle. Nous comparerons alors le comportement observé aux résultats expérimentaux obtenus *in vitro*.

V.1. MATERIEL ET METHODES

C'est le langage Java qui a été utilisé. Les tests ont été réalisés sur huit machines. Quatre sont équipées de processeurs AMD, quatre sont équipées de processeurs Intel. Les quatre machines AMD (LT76, LT77, LT83, LT84) présentent la même configuration matérielle :

- AMD Athlon 64x2 (Brisbane, 2,5 GHz, 2 Coeurs, 2 Threads)
- 2 Go de RAM (DDR2 400 MHz).

Les machines à base de processeur Intel sont toutes différentes.

- PC1 :
 - o Intel Core 2 Duo E8400 (Wolfdale, 3,0 GHz, 2 Coeurs, 2 Threads) ;
 - o 4 Go de RAM (DDR2 400 MHz).
- PC2 :
 - o Intel Core I5 750 (Lynnfield, 2,6 GHz, 4 Coeurs, 4 Threads) ;
 - o 12 Go de RAM (DDR3 1066MHz).
- PC3 :
 - o Intel Core I3 2100 (SandyBridge, 3,1 GHz, 2 Coeurs, 4 Threads) ;
 - o 8 Go de RAM (DDR3 1333MHz).
- PC4 :
 - o Intel Core I5 2410M (SandyBridge, 2,2 GHz, 2 Coeurs, 4 Threads) ;
 - o 8 Go de RAM (DDR3 1333 MHz).

Pour tous les tests décrits ci-après, la durée de la simulation a été mesurée pour 1000 itérations (soit 200 générations avec les paramètres arbitraires utilisés pour la validation du modèle). Les durées indiquées sont des moyennes réalisées sur 30 répliques. Les durées d'exécution ont d'abord été comparées selon la dimension de la matrice. Les tests ont tous été réalisés avec des matrices carrées de dimension 5, 11, 25, 50, 100, 250, 500 et 1000.

Les durées moyennes obtenues avec PC1 :

Taille de la matrice	5	11	25	50	100	250	500	1000
Temps d'exécution (en minutes)	0,17	0,35	0,48	0,57	1,18	6,15	24,70	119,04

Pour comparer les différentes machines, la matrice utilisée est de taille 500x500, le voisinage exploité pour la simulation est le voisinage de Moore de rayon 1.

Machine	PC1	PC2	PC3	PC4	LT76	LT77	LT83	LT84
Temps d'exécution (en minutes)	24,70	14,37	15,18	18,49	50,19	50,13	55,59	50,22

Les processeurs Intel sont plus performants sur cet exercice (environ 15 minutes sur PC1 ou PC2 contre 50 à 55 minutes pour les machines AMD). La fréquence du processeur a un effet net sur la durée de la simulation, ainsi les processeurs plus anciens mais cadencés à une fréquence plus élevée, ou, à génération équivalente, un processeur de PC de bureau cadencé à une fréquence plus élevée qu'un processeur d'ordinateur portable sont plus véloce sur l'exercice que des processeurs fonctionnant à une fréquence plus basse (une quinzaine de minutes avec PC2 et PC3 contre 18 minutes avec PC4).

Il est difficile de généraliser sur la base de nos observations au vu des différences de configuration des machines considérées, il aurait fallu disposer de plusieurs machines partageant la même configuration pour chacun des processeurs Intel décrits, au même titre que nous disposions de plusieurs machines AMD présentant les mêmes caractéristiques.

La quantité de mémoire vive qui peut être allouée à la machine virtuelle Java limite la vitesse d'exécution. Cette limite est liée à un problème de gestion de la mémoire de la machine virtuelle Java. Lorsqu'on enchaîne plusieurs répliques, la mémoire de la machine virtuelle n'est pas libérée correctement, et si elle est limitée à 1 Go, elle finit par « saturer », ce qui ralentit la simulation. Lorsque l'on alloue 2 Go et plus à la machine virtuelle sur les machines qui le permettent, nous n'avons pas observé cette saturation. L'engorgement de la mémoire de la machine virtuelle Java se fait sentir d'autant plus vite que le voisinage exploité dans la simulation est grand. Il se manifeste après cinq répliques avec un voisinage de Moore de rayon 3 alors qu'il ne se ressent qu'après douze répliques avec un voisinage de Moore de rayon 1. Avec 2 Go de mémoire allouée à la machine virtuelle Java, aucun effet n'a été visible même après cent répliques.

L'impact du rayon du voisinage utilisé est important (c'est toujours la même matrice test de taille 500x500 qui est utilisée), par exemple avec PC1 on obtient en moyenne les durées :

Rayon du voisinage de Moore	Moore R1	Moore R2	Moore R3
Temps d'exécution (en minutes)	24,70	46,88	77,47

La plateforme a été développée en utilisant le langage Java et la plateforme de développement NetBeans. Les images présentées dans le manuscrit sont des images générées à partir de fichiers de sorties à un instant choisi, c'est-à-dire que les données de l'état de l'automate cellulaire sont écrites sous forme d'un tableau de chiffres qui est converti en image ou exploité pour réaliser des graphiques. L'interface actuelle qui permet la visualisation en temps réel au cours de la simulation est réalisée à l'aide de la bibliothèque JChart2D avec une applet utilisant la bibliothèque awt.

V.2. PRECISION SUR LE FONCTIONNEMENT DE L'AUTOMATE CELLULAIRE

Avec le modèle proposé, le comportement de l'automate ne présente pas de différence que l'on utilise un processus de mise à jour synchrone (« UpdateAll () ») ou asynchrone (« UpdateAllAsynchrone () »), par contre l'ordre de mise à jour des mailles de l'automate a une forte influence sur son comportement. Si initialement nous avons opté pour une mise à jour des mailles dans l'ordre lexicographique, c'est-à-dire dans l'ordre de leur position dans la matrice, nous avons observé un biais important. Ce processus de mise à jour se traduit par une croissance orientée vers le coin dans lequel a commencé la mise à jour. Par exemple, si que l'ordre de mise à jour suit la matrice ligne par ligne, en commençant par la première cellule en haut à gauche, la population forme une pointe orientée vers le coin supérieur gauche (Figure V.1).



Figure V.1 : Lorsque le processus de mise à jour traite les mailles dans l'ordre lexicographique, on observe un biais dans la croissance qui est orientée vers la première maille, ici située dans le coin supérieur gauche de la matrice. En blanc, la population cellulaire.

Nous avons dû implémenter un processus de mise à jour aléatoire des mailles de l'automate. Nous avons testé deux façons de procéder. Dans un premier temps nous avons créé une séquence de mise à jour en début de simulation, et cette même séquence est répétée tout au long de la simulation. Dans un second temps nous avons créé une nouvelle séquence de mise à jour à chaque itération de la simulation. La seconde méthode est beaucoup plus gourmande en ressource et en temps de calcul pour un résultat qui n'est pas significativement différent. Aussi nous sommes nous satisfaits de la première méthode. Par la suite, lorsque nous parlerons de mise à jour de l'automate, il sera donc question de mise à jour dans un ordre aléatoire choisi en début de simulation et réutilisé ensuite à chaque itération de la simulation. Ceci corrige le biais lié à la mise à jour lexicographique et produit une croissance beaucoup plus homogène dans l'espace de la matrice (Figure V.2).



Figure V.2 : Lorsque le processus de l'automate cellulaire fait appel à un ordre de traitement des mailles aléatoire, la croissance résultante est homogène dans l'espace de la matrice. En blanc, la population cellulaire.

V.3. REPRODUCTIBILITE

Avec cette mise à jour selon un ordre aléatoire fixé au début de la simulation nous avons réalisé 1000 réplifications en utilisant une matrice de taille 100x100 avec 250 itérations. Sur chaque réplification, on compte à l'issue de la simulation le nombre de cellules générées. Avec un ordre de mise à jour des mailles calculé au début de chaque simulation, nous obtenons un effectif minimum de 6314 cellules et un effectif maximum de 6692 cellules avec une moyenne à 6478,4 cellules. Pour comparaison, avec un ordre de mise à jour calculé au début de chaque itération pour chaque simulation, les valeurs sont respectivement 5983, 6323 et 6151,3 pour les effectifs minimum maximum, et moyen. Dans le premier cas, l'écart type représente 1,4 % de l'effectif moyen contre 1,1 % dans le second cas (Tableau V.1).

	ORDRE DE MISE A JOUR CALCULE AU DEBUT DE LA SIMULATION	ORDRE DE MISE A JOUR CALCULE A CHAQUE ITERATION
EFFECTIF MINIMUM	6314	5983
EFFECTIF MAXIMUM	6692	6323
EFFECTIF MOYEN	6478,4	6151,3
ÉCART TYPE	92,9	72,3

Tableau V.1 : Bilan des effectifs cellulaires obtenus au cours de 1000 réplifications du modèle selon l'ordre de mise à jour utilisé.

En reportant le nombre de fois où chaque effectif a été obtenu et en les normalisant par rapport la moyenne et l'écart type calculés sur ces 1000 réplifications on obtient le graphe suivant (Figure V.3). Avec ce résultat nous confirmons que le modèle produit des résultats reproductibles qui présentent un écart-type acceptable.

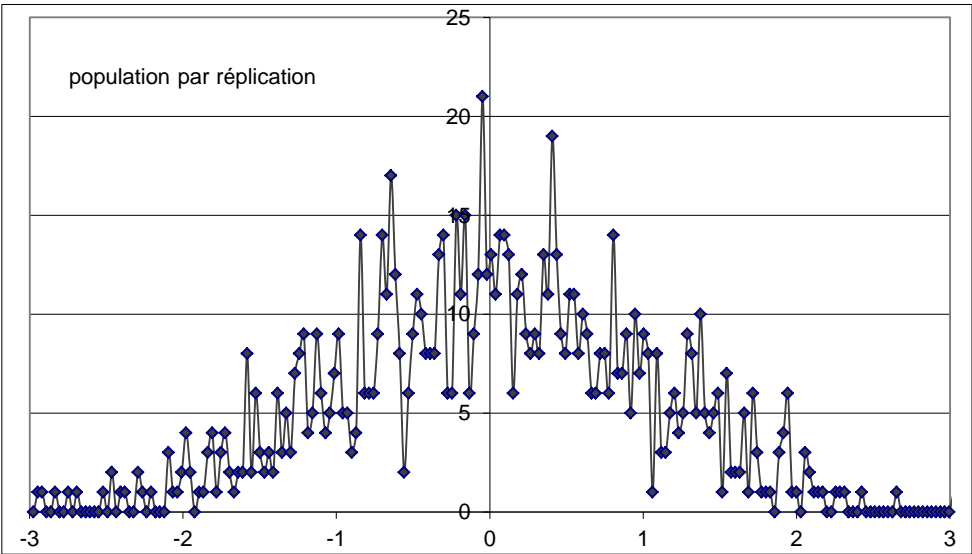


Figure V.3 : Distribution des effectifs obtenus en réalisant 1000 simulations du modèle sur la plateforme dans une matrice de 100x100, l'ordre de mise à jour des mailles est calculé au début de chaque simulation.

V.4. VALIDATION DU COMPORTEMENT DU MODELE ET DE LA ROBUSTESSE DES RESULTATS

Le modèle implémenté dans la plateforme a montré des résultats reproductibles. Pour l'instant nous avons surveillé son comportement général et contrôlé son comportement selon les processus de mise à jour utilisés. Maintenant, nous allons vérifier plus en détail que le modèle présente les bonnes caractéristiques, dans son comportement général, et dans le comportement individuel de chaque cellule.

V.4.1. JEUX D'ESSAIS

V.4.1.1. PARAMETRES

Pour vérifier le comportement du modèle, toutes les cellules sont initialisées avec les mêmes valeurs de paramètres cohérents, mais sans signification biologique. La variabilité interindividuelle est nulle. Ainsi toutes les cellules, de la première génération ou des générations suivantes, sont identiques, avec les mêmes valeurs de paramètres et donc, le même comportement.

Pour faciliter la vérification nous plaçons ces cellules dans un contexte qui modélise un milieu solide, c'est-à-dire qu'il n'y a pas d'agitation du milieu ; une cellule conserve la position où elle est créée tout au long de la simulation.

V.4.1.1.1. PARAMETRES DES CELLULES

- $\text{MAX_CONSUMPTION} = 1$;
- $\text{REQUIREMENTS} = 1$;
- $\text{MIN_CYCLE_LENGTH} = 5$;
- $\text{MAX_UPTAKE} = 1$;
- $\text{MAX_OUTTAKE} = 1$;
- $\text{RELEASED_COMPOUNDS} = 0,25$;
- $\text{max_consumption}_i = (\text{MAX_CONSUMPTION}) * 1$
- $\text{requirements}_i = (\text{REQUIREMENTS}) * 1$
- $\text{min_cycle_length}_i = (\text{MIN_CYCLE_LENGTH}) * 1$
- $\text{max_uptake}_i = (\text{MAX_UPTAKE}) * 1$
- $\text{max_outtake}_i = (\text{MAX_OUTTAKE}) * 1$

V.4.1.1.2. VARIABLES DES CELLULES

- $\text{cycle}_i = 0$;
- $\text{El}_i = 1$;
- $\text{readiness}_i = \text{false}$.

V.4.1.1.3. PARAMETRES UTILISES DANS LES FONCTIONS

- growthFactor = 2 (*utilisé dans dividingCellFunction*);
- environmentFactor = 1 (*utilisé dans dividingCellFunction et CellFunction*);
- degradationRate = 0,15 (*utilisé dans dividingCellFunction et CellFunction*).

V.4.1.1.4. CAMPAGNE DE TESTS

Les différents essais menés ont été réalisés sur une matrice carrée. Le pas de temps de la simulation est estimé à 5 minutes. Différentes configurations ont été testées, en faisant varier la taille de la matrice, la quantité de nutriments disponibles dans l'environnement, la taille du voisinage utilisée, le nombre et la disposition des cellules initiales.

Nous ne présenterons par la suite que les essais réalisés sur une matrice de dimension 500x500, avec une quantité initiale de nutriments de 30 millions d'unités, sur une durée de 1000 pas de temps notés timeStep.

Dans ces mêmes conditions, quatre dispositions initiales des cellules ont été testées, en exploitant pour chacune 3 tailles de voisinage (voisinage de Moore de rayon 1, de rayon 2 et de rayon 3, voir Figure IV.2) notés R1, R2 et R3.

Pour rappel, le voisinage sert lors du processus de division cellulaire à identifier les mailles libres à proximité de la cellule qui se divise et qui seront éligibles pour accueillir la nouvelle cellule fille lorsque la notion de superindividu n'est pas utilisée.

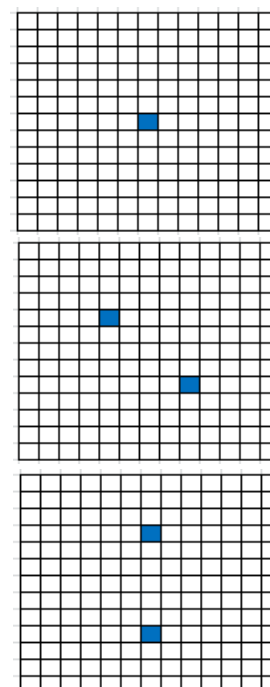
Les quatre dispositions initiales sont :

Cas 1 : 1 cellule seule au centre de la matrice (poplnit = 1*1)

Cas 2 : 1 population de 162 cellules au centre de la matrice (poplnit = 1*162)

Cas 3 : 2 populations de 81 cellules chacune, soit 162 cellules au total, disposées sur la diagonale de la matrice (poplnit = 2*81)

Cas 4 : 2 populations de 81 cellules chacune, soit 162 cellules au total, disposées sur la médiane verticale de la matrice (poplnit = 2*81)



Ceci nous permet de comparer la croissance observée à partir d'un foyer : une cellule ou d'une population (cas 1 et 2). Nous vérifions le comportement des populations en cas de

« collision » (cas 3 et 4) et enfin nous comparons la croissance observée à partir d'un effectif total identique, selon qu'il y a 1 ou plusieurs populations au départ (cas 2, 3 et 4).

Remarque : pour les cas 3 et 4, la distance séparant les deux populations est comparable, aux arrondis près à cause de la discrétisation de l'espace, il y a le même nombre de mailles séparant les centres des 2 foyers disposés en diagonale du cas 3 qu'il y a de mailles séparant les centres des 2 foyers disposés sur la médiane du cas 4.

Des campagnes de tests ont été lancées sur plusieurs machines afin de vérifier en même temps l'influence du matériel. 10 répliquations de l'automate ont été lancées pour chacune des configurations initiales avec chacun des 3 rayons de voisinage sur chaque machine, soit un total de 60 essais pour chaque configuration et initiale et pour chaque rayon de voisinage afin de vérifier la reproductibilité des résultats et valider le comportement du modèle avec ces paramètres simples.

V.4.1.2. CROISSANCE OBSERVEE

Dans tous les cas considérés, le comportement observé est similaire quel que soit le rayon de voisinage utilisé. La différence tient à la croissance qui est facilitée lorsque le rayon de voisinage considéré augmente, ce qui se traduit par une croissance plus rapide.

V.4.1.2.1. CROISSANCE DEPUIS UN FOYER

Avec les cas 1 et 2, les courbes sont similaires et aboutissent à un effectif final comparable. La différence de population initiale se reporte sur le temps nécessaire à atteindre le maximum de la population (Figure V.5 et Figure V.7).

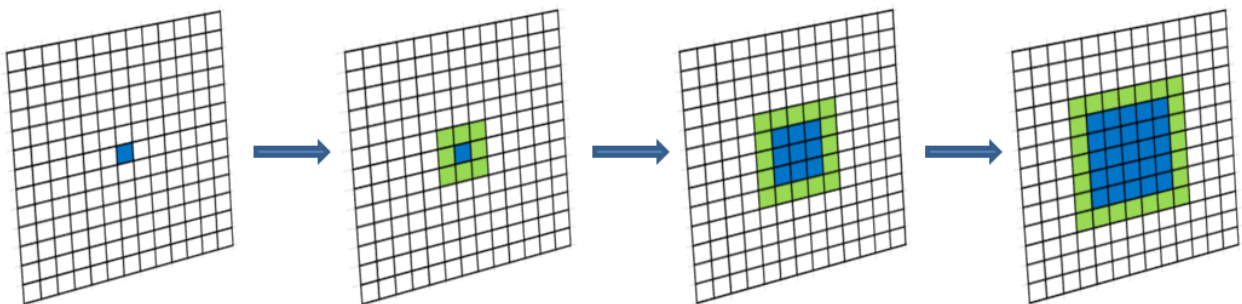


Figure V.4 : Schéma de la croissance de la population dans les cas 1 et 2.

Le temps mis à atteindre le plus grand nombre de cellules avec le cas 1 est en effet plus long que celui observé avec le cas 2 d'un nombre d'itérations correspondant approximativement au nombre de pas de temps nécessaires à obtenir une population de 162 cellules à partir d'une unique cellule initiale.

Par exemple, si l'on regarde ce qu'il se passe en utilisant le voisinage de rayon 2. On voit qu'avec le cas 1, à partir d'une seule cellule initiale, la taille maximale de la population, 136587 en moyenne, est obtenue après 687 itérations en moyenne, alors qu'avec le cas 2, à partir d'une population initiale de 162 cellules, le maximum, 137831 en moyenne, est atteint après 642 pas de simulation en moyenne. Le cas 2 atteint donc son maximum 45 pas de

temps plus tôt que le cas 1. Or si l'on regarde les données de la croissance de l'effectif du cas 1, on observe qu'en moyenne, on atteint une population de 119 à l'itération 42 et une population de 194 après 48 pas de temps. Si l'on fait la moyenne de ces pas de temps et populations, on obtient une population moyenne de 156 après 45 pas de temps.

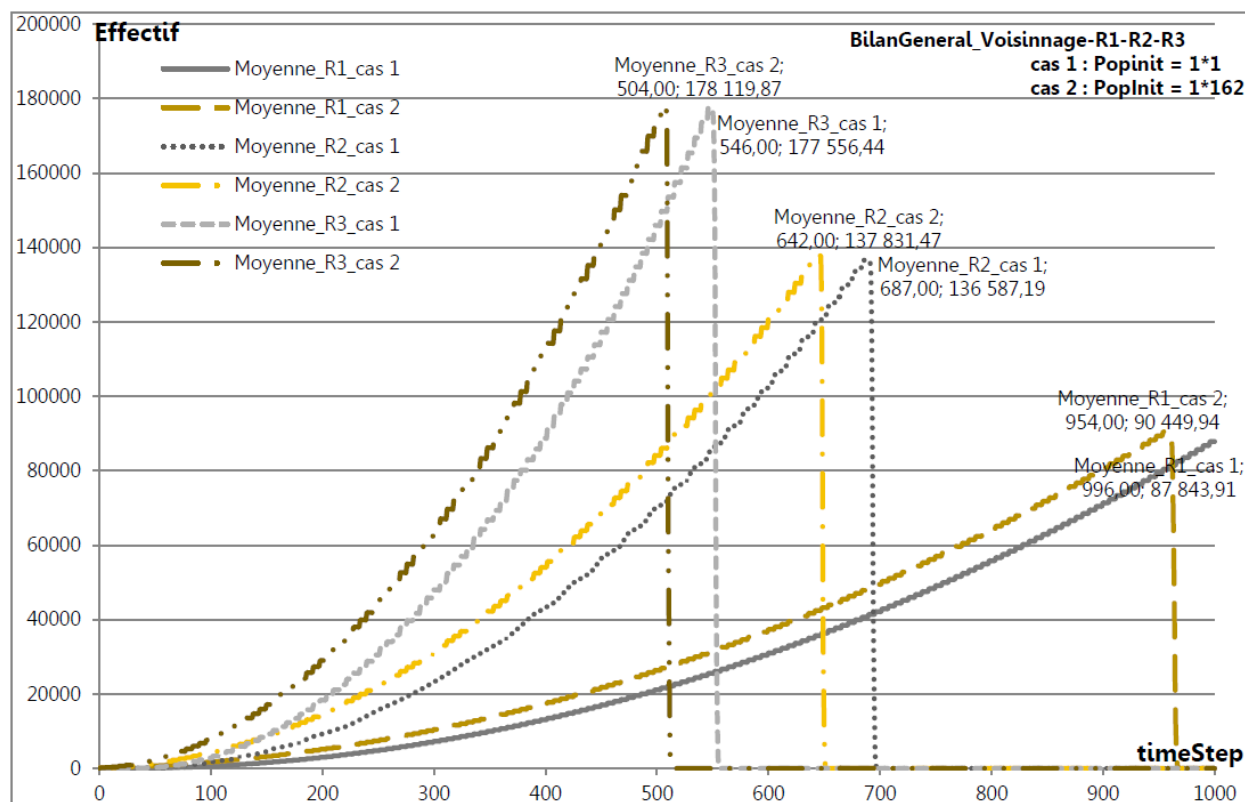


Figure V.5 : Courbes de croissance moyennes obtenues avec les Cas 1 et 2 exploités chacun avec les voisinages de Moore de rayon 1 (R1), de rayon 2 (R2) et de rayon 3 (R3).

Ce comportement est bien celui attendu, on part d'une population initiale plus importante et on atteint par conséquent rapidement l'épuisement des nutriments du milieu. Cependant, la croissance obtenue est trop lente par rapport à ce que l'on attend. Les cellules qui disposent de ressources suffisantes progressent dans leur cycle cellulaire, puis se divisent. Avec une valeur de MIN_CYCLE_LENGTH à 5, on s'attendrait à voir chaque cellule se diviser tous les 5 pas de temps, et donc à voir la population doubler tous les 5 pas de temps. Ce qui serait le comportement d'une population qui a une croissance exponentielle à taux constant.

Si on analyse les données obtenues selon une échelle logarithmique, on constate que le taux de croissance diminue dans le temps. La croissance est freinée au fur et à mesure que la population se développe. C'est normal en examinant les règles. Seules les cellules qui disposent d'espace libre à proximité peuvent se diviser. La croissance n'est donc pas fonction de la population, mais de la fraction de la population, qui par son voisinage accède à des mailles vides.

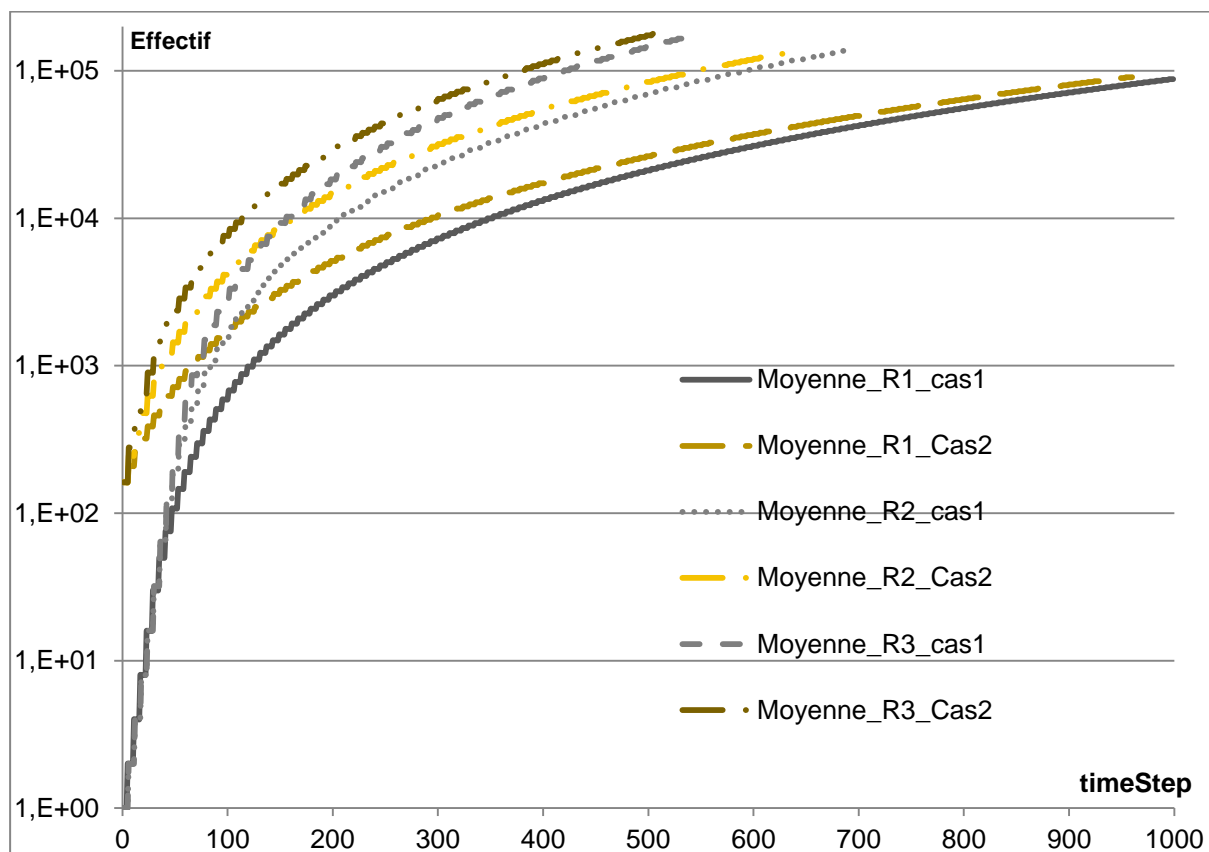


Figure V.6 : Croissance comparée en échelle logarithmique des cas 1, 2 et 3 en utilisant les voisinages de Moore de rayon 1, 2 et 3 pour chaque cas.

V.4.1.2.2. CROISSANCE A PARTIR DE PLUSIEURS FOYERS

Avec les cas 3 et 4, on observe dans un premier temps une croissance comparable. Lorsque les 2 populations entrent en collision, on a un infléchissement de la courbe de croissance lié au nombre de cellules qui ne disposent plus d'espace pour se diviser.

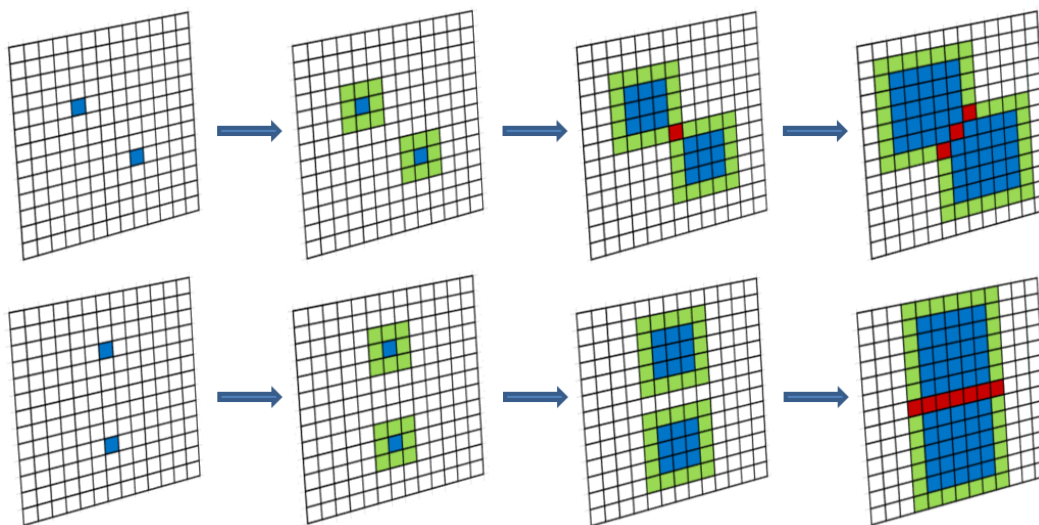


Figure V.7 : Schéma de la croissance dans les cas 3 et 4. La zone de collision des deux populations est mise en évidence.

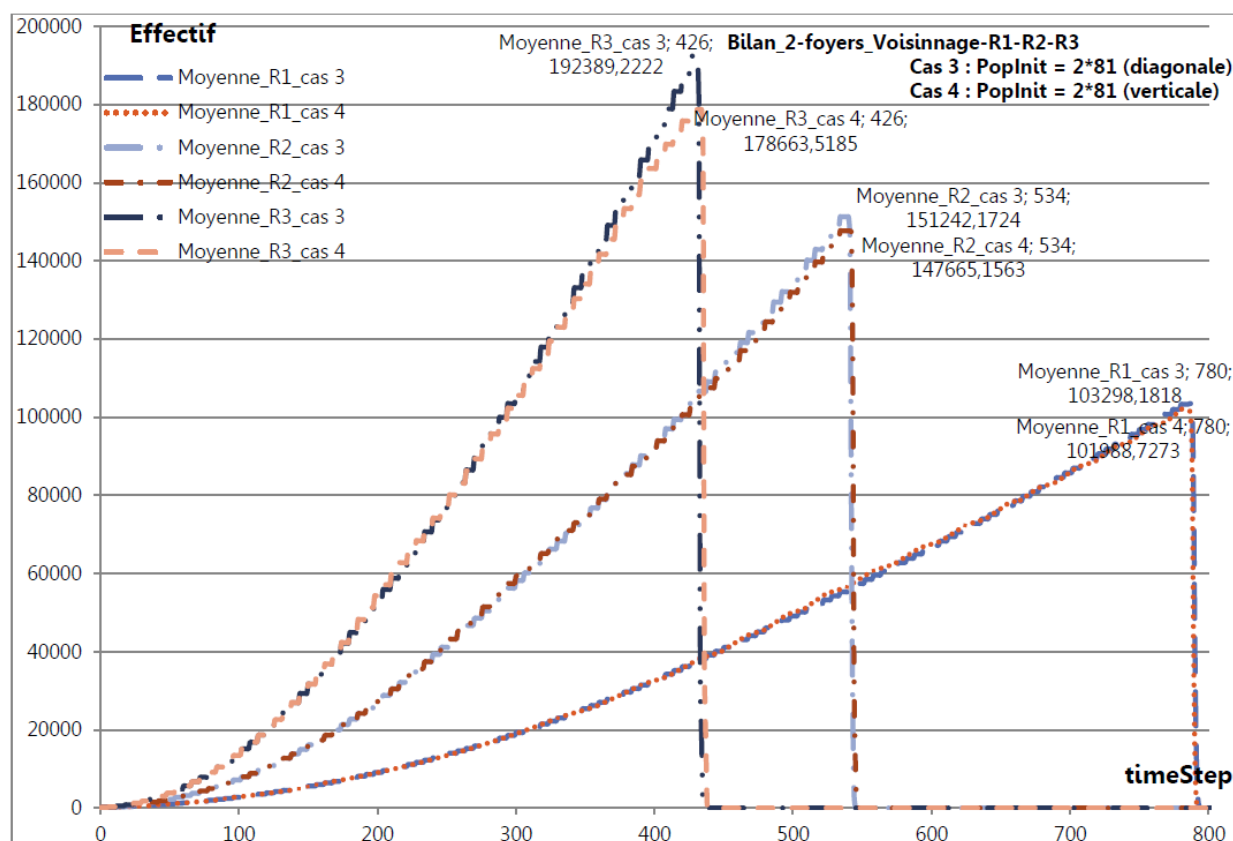


Figure V.8 : Courbes de croissance moyennes obtenues avec les Cas 3 et 4 exploités chacun avec les voisinages de Moore de rayon 1 (R1), de rayon 2 (R2) et de rayon 3 (R3).

	MOYENNE DES POPULATIONS OBTENUES	POPULATION MINIMALE OBTENUE	POPULATION MAXIMALE OBTENUE	ECART TYPE (INDIVIDUS)	ECART TYPE RELATIF (%)
CAS 1 R1	87788,93	87244,00	88587,00	323,70	0,37
CAS 1 R2	136701,78	136130,00	138292,00	422,51	0,31
CAS 1 R3	177556,44	175018,00	178408,00	643,16	0,36
CAS 2 R1	90449,94	90033,00	90961,00	221,03	0,24
CAS 2 R2	137831,47	137321,00	138380,00	249,14	0,18
CAS 2 R3	178119,87	177007,00	178489,00	266,59	0,15
CAS 3 R1	103418,38	103069,00	104209,00	222,86	0,22
CAS 3 R2	137831,47	137321,00	138380,00	249,14	0,18
CAS 3 R3	192389,22	191727,00	193090,00	288,19	0,15
CAS 4 R1	101988,73	101510,00	102523,00	190,07	0,19
CAS 4 R2	147665,16	147141,00	148315,00	237,68	0,16
CAS 4 R3	178663,52	178110,00	179132,00	197,92	0,11

Tableau V.2 : Récapitulatif des résultats obtenus par l'étude des cas 1, 2, 3 et 4 à l'aide du voisinage de Moore de rayon 1, 2 et 3.

Du fait du maillage de l'automate et du voisinage utilisé, voisinage de Moore, la zone de contact entre les 2 populations est plus importante dans le cas 4, où elles sont disposées sur

la médiane de la matrice, que dans le cas 3, où elles sont disposées sur la diagonale de la matrice (Figure V.7). Cela se traduit par un infléchissement plus important de la courbe de croissance dans le cas 4 que dans le cas 3. Plus le rayon du voisinage utilisé est grand, plus la croissance est rapide et plus cette différence de croissance entre ces deux cas est marquée (Figure V.8).

Ici encore ce comportement est celui que l'on s'attendait à avoir. En effet, la zone de contact étant plus importante dans le cas 4, l'infléchissement de la courbe est plus marqué et correspond à nos attentes.

V.4.2. BILAN DES RESULTATS SANS EXPLOITER LE SUPERINDIVIDU

Dans les 4 cas observés la quantité de nutriments constitue l'élément limitant la croissance.

Dans tous les cas, et quel que soit le rayon de voisinage employé, les résultats présentent une faible variabilité et sont reproductibles. L'aléatoire intervient qui dans l'ordre de mise à jour des cellules, et lors de la division, dans le choix de la maille de destination de la cellule fille ne produit pas de biais dans le comportement du modèle.

Notons que le choix aléatoire de la maille de destination de la cellule fille lors de la division a un effet plus marqué lorsque le rayon de voisinage employé augmente. Ceci s'explique par le fait que les cellules sont plus souvent gênées et ne se divisent plus lorsque le rayon de voisinage de Moore choisi est plus faible. Un voisinage plus étendu autorise une plus grande dispersion des cellules dans la matrice lors de leur division, elles ont donc tendance à moins se gêner les unes les autres et peuvent ainsi continuer à se diviser plus longtemps (Tableau V.2).

La croissance bactérienne que l'on observe *in vitro* est exponentielle. Avec les résultats décrits précédemment, nous sommes loin d'afficher une croissance aussi soutenue.

Nous avons identifié que la croissance ne se fait que par les cellules à la périphérie. Cela révèle que nos hypothèses initiales :

- Ne considérer qu'un individu par maille
- La division d'une cellule se fait dans le système d'interaction de l'automate cellulaire

Contraignent la croissance qui est fortement limitée : les cellules qui ne disposent plus d'espace libre dans leur voisinage ne peuvent plus se diviser. Ainsi, lorsque l'on augmente le rayon du voisinage utilisé, la croissance est plus rapide (Tableau V.2). On peut extrapoler et supposer que si le voisinage utilisé tend vers l'infini, ce biais diminuera et la croissance simulée tendra vers la croissance exponentielle observée expérimentalement.

Pour le vérifier nous avons mené quelques essais à partir du Cas 1, d'abord en considérant un milieu agité dans lequel toutes les cellules sont redistribuées aléatoirement dans l'espace à chaque itération avec un voisinage de Moore de rayon 1, puis en considérant un milieu solide mais avec l'ensemble de l'espace disponible comme voisinage.

La croissance qui s'observe alors est exponentielle jusqu'à saturation de la matrice (Figure V.9 ; 250 000 individus avec la matrice utilisée).

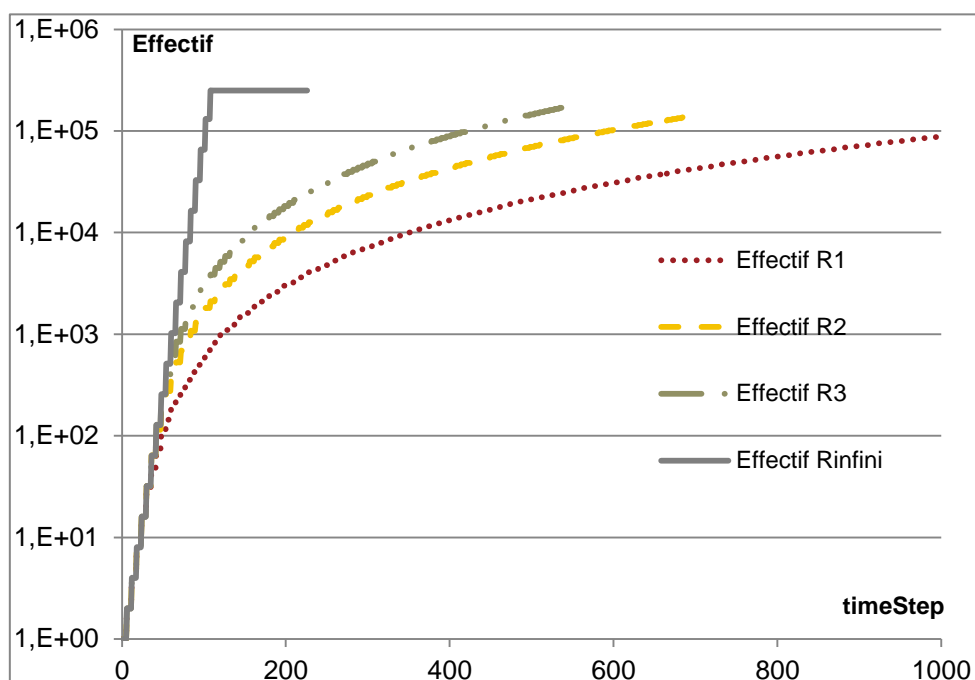


Figure V.9 : Comparaison, en échelle logarithmique, de la croissance obtenue dans le Cas 1 selon le rayon du voisinage de Moore employé.

Nous savons obtenir une croissance exponentielle, mais la taille de la population obtenue reste limitée par l'espace modélisé. Pour nous libérer de cette contrainte liée à l'espace, nous nous orientons vers le système des superindividus.

V.4.3. RESULTATS AVEC UTILISATION DU SUPERINDIVIDU

L'implémentation s'accompagne de quelques modifications supplémentaires. Du fait de la topologie de l'espace de l'automate cellulaire, une cellule qui se divise mais dispose de place autour d'elle s'initialise et conserve sa valeur de cellulesParMaille à 1, et la cellule fille générée dans une maille voisine obtient une valeur de cellulesParMaille de 1. Lorsque c'est une cellule qui n'a plus d'espace libre autour d'elle qui se multiplie, la valeur de cellulesParMaille est doublée. On considère ainsi que toutes les cellules représentées par un superindividu se divisent en même temps.

Cette modification lève la contrainte de l'espace sur la croissance et permet d'obtenir une courbe plus en adéquation avec ce que donne la courbe théorique décrite dans INDISIM par exemple (Figure V.10).

Ceci dit, si le biais imposé par l'espace est supprimé par cette modification, elle introduit un nouveau biais. S'il peut exister une variabilité entre les cellules occupant deux mailles voisines, avec la notion de superindividu, chaque maille est occupée par une « super-cellule », tous les individus présents dans un de ces superindividus partagent les mêmes paramètres. C'est-à-dire que tous les individus qui forment un superindividu sont forcément

synchronisés et partagent exactement les mêmes valeurs à tous leurs paramètres. L'exemple le plus parlant sera celui de l'état interne qui a une valeur unique que le nombre d'individus soit 1 ou plus. Cela signifie que tous les individus du superindividu sont affectés de la même manière et en même temps par l'apparition d'une carence.

Cela se traduit lorsqu'on observe la croissance par une mort plus rapide des cellules du centre de la population. En effet, lorsque les nutriments manquent, on évalue la carence en fonction du besoin total du superindividu. Plus cellulesParMaille a une valeur importante, et plus le manque est important. Comme le devenir de l'état interne est calculé en fonction du manque, l'état interne est d'autant plus diminué que le manque est important, et donc, que cellulesParMaille a une valeur élevée.

Les premières cellules à être limitées par leur espace sont celles qui ont la plus grande valeur de cellulesParMaille (elle double à chaque cycle de division), ce seront donc elles qui verront leur état interne décroître le plus vite lorsque les nutriments viendront à manquer, et donc ce sont elles qui mourront le plus vite. Cela se traduit par la disparition de la population du centre vers la périphérie alors que cette disparition était homogène avant ce changement.

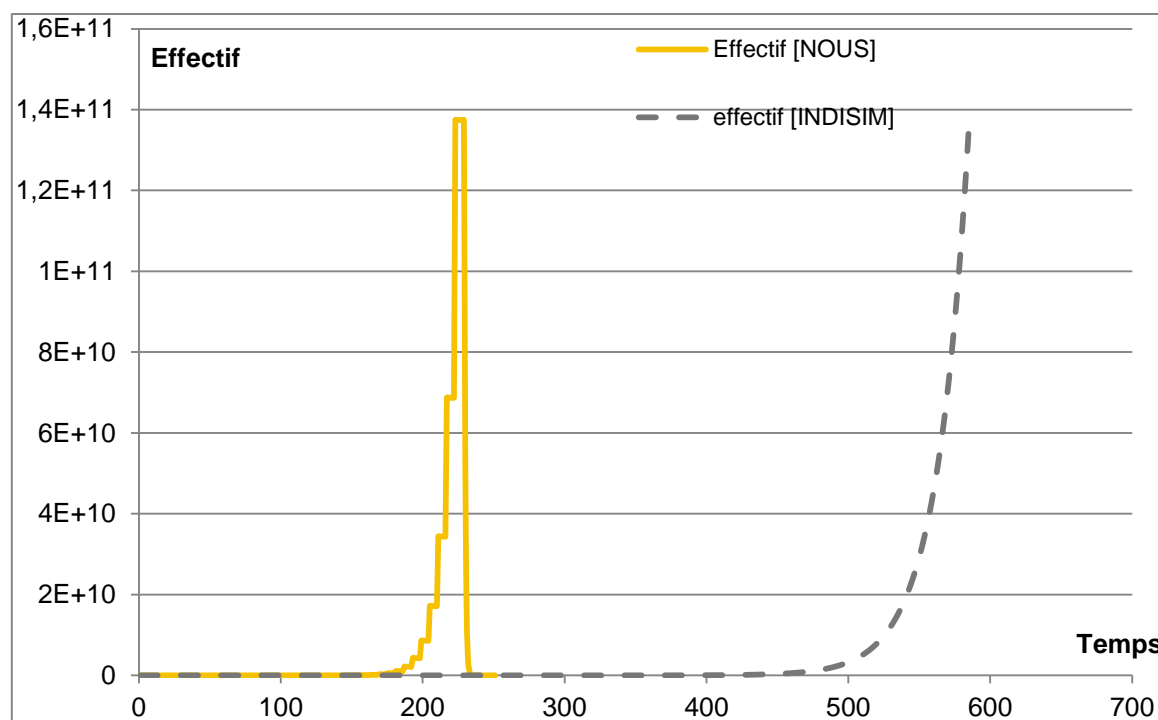


Figure V.10 : Comparaison de la courbe de croissance fournie par INDISIM à celle que nous obtenons.

Dernier point qui est biaisé par ce changement, la valeur de cellulesParMaille entre 2 cellules adjacentes peut être très importante. Puisqu'actuellement, cette valeur double si la cellule qui se divise n'a plus de place autour d'elle, il arrive que certaines cellules aient une valeur de cellulesParMaille de plusieurs millions alors que leurs voisines immédiates n'ont une valeur de ce paramètre que de quelques centaines de milliers.

V.5. UTILISATION DES PARAMETRES ISSUS D'ETUDES BIOLOGIQUES

V.5.1. IMPLEMENTATION

Nous avons calculé plus tôt qu'une bactérie consomme $1,636 \times 10^9$ molécules de glucose par heure, ce qui revient à dire qu'elle consomme $2,73 \times 10^7$ molécules de glucose par minute.

Pour simplifier l'écriture, nous avons pris l'hypothèse que 1 unité de nutriments dans le modèle correspondait à 10^7 molécules de glucose.

Avec la possibilité de mettre plusieurs bactéries dans une maille, l'estimation du volume modélisé par rapport au nombre maximum de cellules pouvant prendre place n'avait plus de sens. Par contre, la concentration du milieu en glucose reste inchangée, elle est toujours de 2g/L soit 11×10^{-3} mol/L, soit 11 mmol/L.

Nous avons vu qu'une unité de DO correspond à 10^9 cellules par millilitre. En milieu riche, on obtient facilement aux alentours de 10 de DO, soit une densité de cellules de 10^{10} cellules/mL.

Partant de ces hypothèses et en prenant pour objectif d'utiliser les paramètres obtenus *in vitro* par C. Nicolas nous avons changé les paramètres par défaut utilisés dans le modèle jusqu'à présent. Nous avons ainsi remplacé la valeur de cycle qui était fixée arbitrairement à 5 par 64, temps de génération décrit dans ses travaux, et les besoins qui avaient été fixés à 1 ont été remplacés par 2,73. Ce faisant, nous avons estimé le pas de temps du modèle à 1 minute.

V.5.2. RESULTATS

Avec un nombre illimité de bactéries le volume modélisé est impossible à estimer, nous avons pris l'hypothèse d'un volume de milieu de 1 mL. Au vu de la concentration de 11 mmol/L et en vertu du nombre d'Avogadro, cela correspond à $1,083 \times 10^{20}$ molécules de glucose. Avec l'hypothèse qu'1 nutriment vaut 10^7 molécules, nous avons décidé d'introduire dans l'environnement $1,083 \times 10^{13}$ unités de nutriments.

L'effectif maximum obtenu est alors de l'ordre de 7×10^{10} bactéries, soit de l'ordre de la densité de cellules obtenue expérimentalement avec un milieu riche.

La courbe de la croissance que l'on observe alors suit la courbe de croissance théorique obtenue avec les paramètres décrits dans la thèse de C. Nicolas. (Figure V.11). L'estimation de la durée modélisée par un pas de temps, une minute est confirmée.

Nous souhaitons cependant nous autoriser de gérer le temps simulé plus finement, en modifiant la durée correspondante à un pas de temps. Par exemple, en envisageant qu'un pas de temps représente 30 secondes pour affiner la modélisation.

Il nous faut alors envisager une manière simple de modifier la durée modélisée par pas de temps, et ajuster en conséquence tous les paramètres qui en découlent. Par exemple, la consommation est fixée à 2,73 nutriments par minute. Avec l'hypothèse d'une minute par pas de temps cela revient à 2,73 nutriments par pas de temps. Pour considérer qu'un pas de temps représente 30 secondes, la consommation doit être corrigée à 1,365 nutriments par pas de temps.

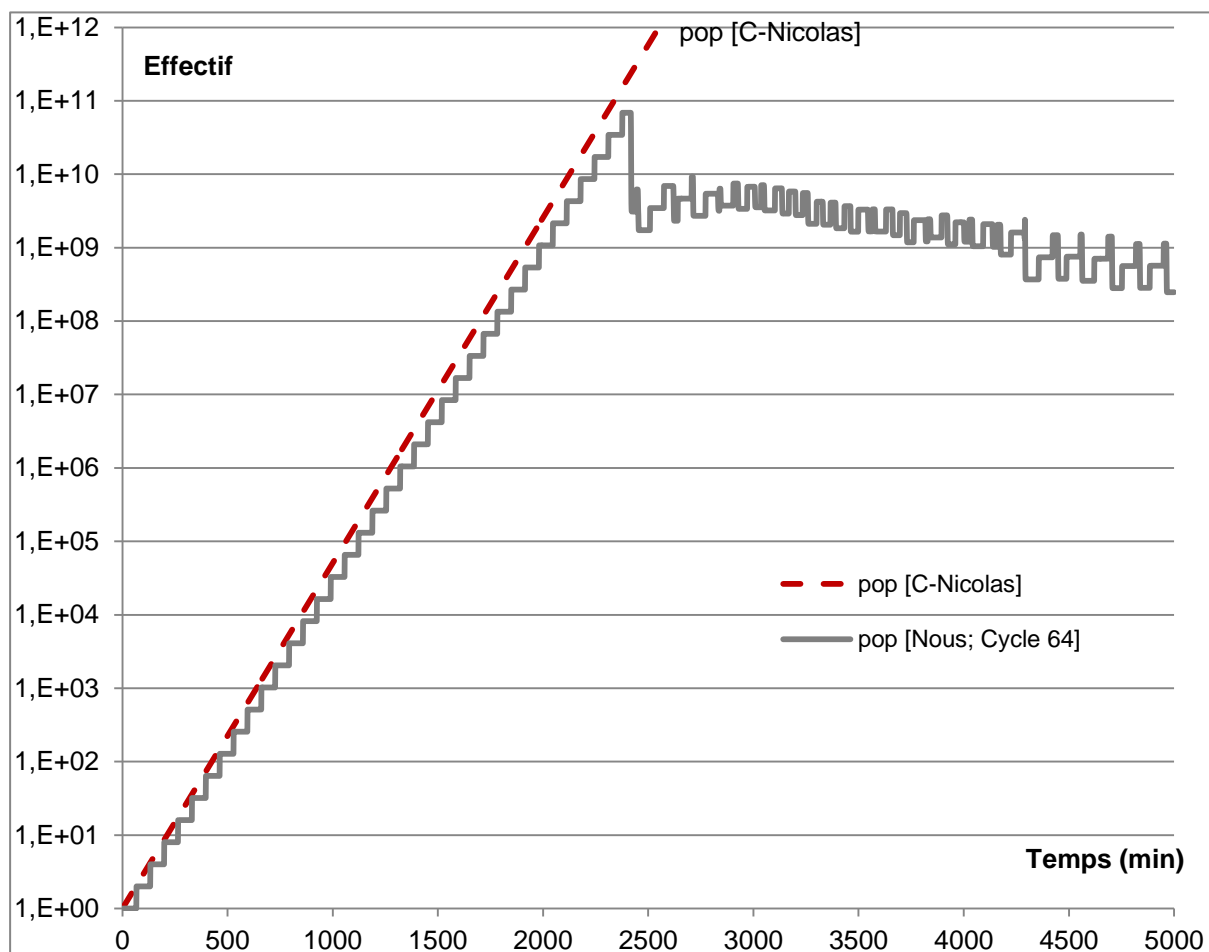


Figure V.11 : Courbes de croissance obtenues en utilisant les paramètres tirés de la thèse de C. Nicolas comparée à la croissance théorique qui permet d'obtenir ces valeurs de paramètres.

Pour faciliter cette correction, nous avons introduit au niveau de la gestion de l'automate cellulaire un nouveau paramètre, `facteurTemps`, qui relie le temps simulé au pas de temps de la simulation.

Les paramètres utilisés deviennent alors :

- `facteurTemps` = 1 ;
- `MAX_CONSUMPTION` = 2,73 x `facteurTemps` ;
- `REQUIREMENTS` = 1 x `facteurTemps` ;
- `MIN_CYCLE_LENGTH` = 91 ;
- `MAX_UPTAKE` = 3 x `facteurTemps` ;
- `MAX_OUTTAKE` = 1 x `facteurTemps` ;
- `RELEASED_COMPOUNDS` = 0,00025.

Les fonctions f_4 , f_5 et f_6 deviennent alors :

- $f_4 : El_i \leftarrow El_i + \text{excédent} * 0,25 * \text{environnement}$;
- $f_5 : El_i \leftarrow El_i - \text{manque} * \text{environnement}$;
- $f_6 : \text{cycle}_i \leftarrow \text{cycle}_i + El_i * \text{facteurTemps}$

Avec plusieurs cellules par maille, le rayon du voisinage utilisé ne change rien à la courbe de croissance obtenue. Par la suite, nous n'avons donc plus employé que le voisinage de Moore de rayon 1 qui est le moins gourmand en temps de calcul.

V.5.3. PHASE STATIONNAIRE

Parallèlement à ce début de paramétrage en fonction de données biologiques, nous avons essayé d'estimer les autres paramètres impliqués : la façon dont la cellule abimée restaure son état interne, la quantité d'éléments nutritifs libérés dans l'environnement lorsqu'une cellule meurt, la façon dont se dégrade l'état interne d'une cellule carencée, etc...

Jouer sur tous ces paramètres permet d'avoir une phase stationnaire après la phase exponentielle de croissance. Cette phase commence dès lors qu'il n'y a plus de nutriments disponibles dans l'environnement. Nous avons trouvé un ensemble de valeurs qui permet d'obtenir un équilibre entre les cellules qui meurent et les cellules qui survivent (Figure V.12).

Cet équilibre est relativement précaire et présente une certaine instabilité. En effet, lorsqu'une cellule meurt elle libère des nutriments dans l'environnement. Par extension, en considérant une cellule comme un superindividu, cette quantité de nutriments libérée est multipliée par le nombre de cellules représentées par le superindividu. Avec cette notion de superindividu, il peut exister une très forte différence du nombre de cellules représentées entre les différentes mailles. Ainsi, selon l'ordre de mise à jour des mailles de l'automate, si les premiers individus à souffrir du manque de nutriments, et donc les premiers à mourir, représentent un grand nombre de cellules, la quantité de nutriments libérés pourra être très importante. Cela se traduira par une reprise intense de la croissance jusqu'à nouvel épuisement de l'environnement. Inversement, si les premiers individus qui meurent représentent peu de cellules, il y aura peu de nutriments libérés, et la croissance ne pourra pas reprendre efficacement.

On a ainsi une oscillation, une population se développe puis meurt, autorisant le développement d'une autre partie de la population, et ainsi de suite. L'aléatoire qui intervient lors de la définition de l'ordre de mise à jour des mailles de l'automate provoquera des oscillations plus ou moins importantes. Ceci est particulièrement marqué en partant d'une cellule initiale unique représentant un seul individu.

La simulation a alors été conduite en partant non plus d'une cellule unique, mais d'une population de 1 000 cellules représentant chacune un individu unique. Deux populations différentes ont été utilisées :

- Un groupe de 1 000 cellules : 1x1k
- 1 000 cellules isolées : 1kx1

La courbe de croissance est similaire, l'effectif maximum est comparable que l'on parte d'une cellule isolée, de 1000 cellules isolées, ou d'un massif de 1 000 cellules.

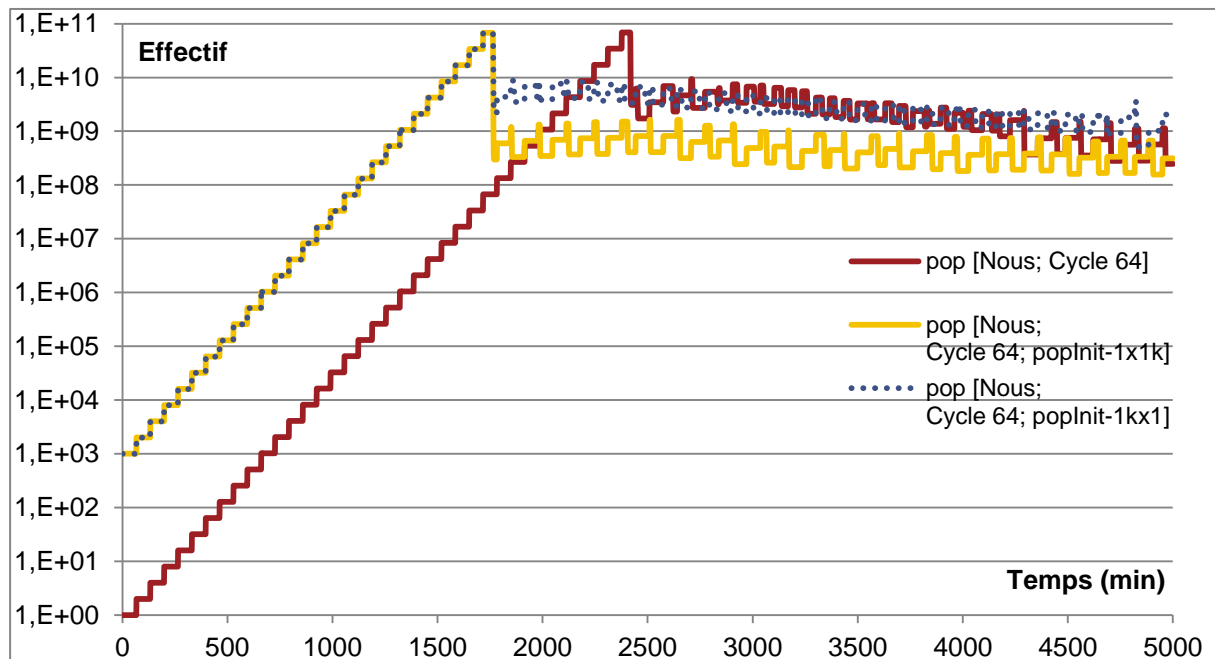


Figure V.12 : Courbes de croissance avec apparition de la phase stationnaire à partir des différentes configurations initiales de la population, avec des paramètres fixes et des cellules synchrones.

Partant de cette vérification de l'aléatoire a été introduit dans les paramètres des cellules. D'abord seulement lors de la première génération, c'est-à-dire à la création de la population (Figure V.13).

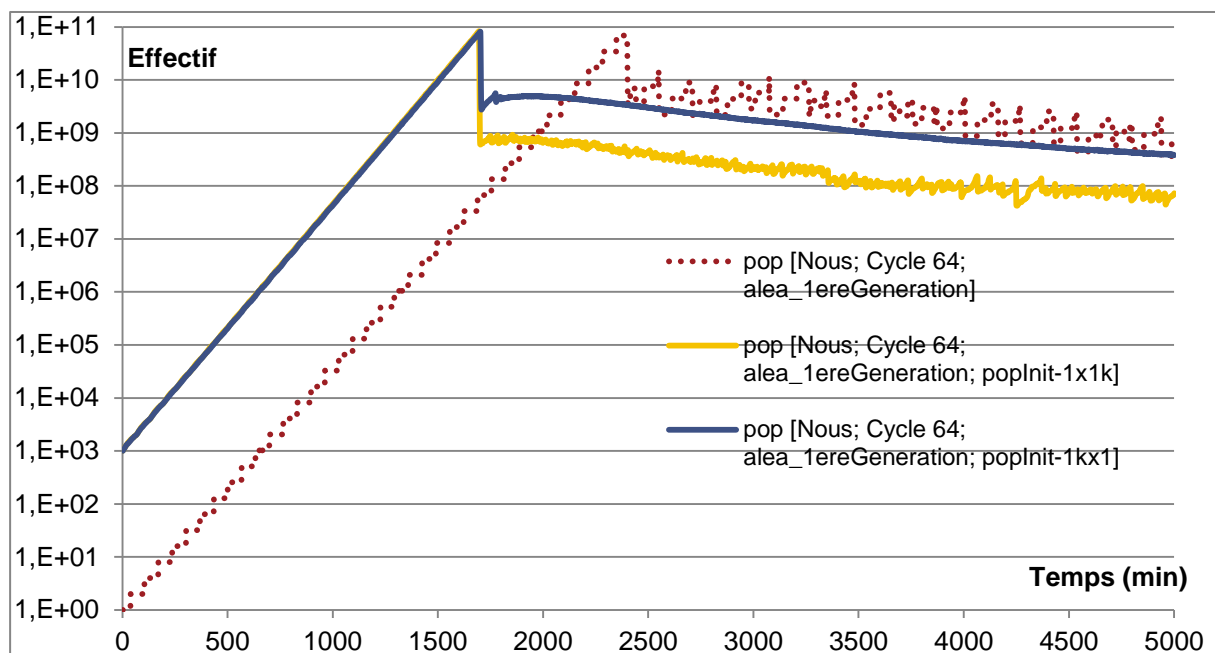


Figure V.13 : Obtention de la phase stationnaire de la croissance à partir des différentes configurations initiales de la population, avec des paramètres générés aléatoirement lors de la création de la génération 0.

Lorsqu'on part d'une cellule unique, la courbe ne présente pas de différence significative par rapport à celle obtenue sans introduire d'aléatoire. La seule différence notable est un décalage dans le temps qui correspond à l'âge de la cellule initiale créée.

Avec une population initiale de 1000 cellules, la courbe obtenue est plus « lisse ». Ceci est lié au fait que chaque cellule créée a un âge aléatoire entre 0 et la longueur de son cycle, si bien qu'une partie de la population pourra se diviser à chaque itération.

Dans un second temps, les paramètres de toutes les cellules ont été générés aléatoirement à chaque division cellulaire (Figure V.14).

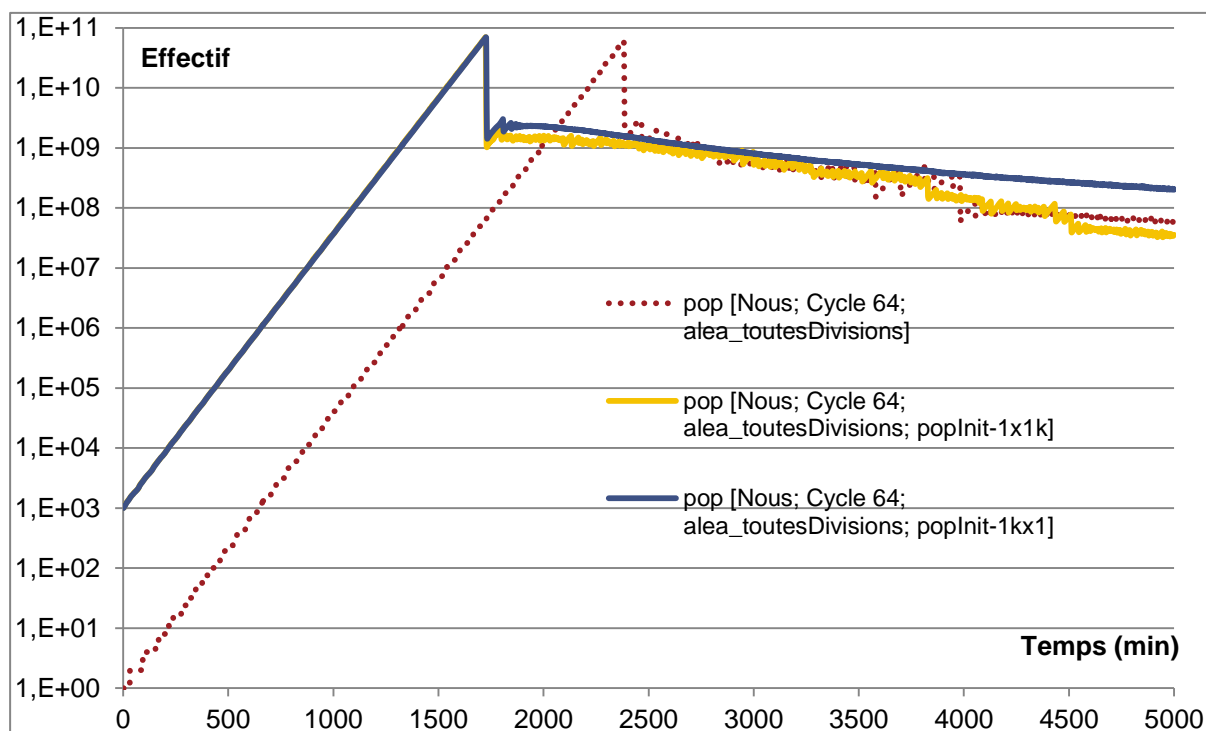


Figure V.14 : Obtention de la phase stationnaire de la croissance à partir des différentes configurations initiales de la population, avec des paramètres générés aléatoirement lors de chaque division cellulaire.

La distribution aléatoire des paramètres propres à chaque cellule par rapport aux valeurs de référence pour l'espèce, la fonction f1, est encore à affiner. Pour les essais décrits ici, voilà ce qui a été testé :

- $\text{max_consumption}_i \in [\text{MAX_CONSUMPTION} - 0,25 ; \text{MAX_CONSUMPTION} + 0,25]$;
- $\text{requirements}_i \in [\text{REQUIREMENTS} - 0,1 ; \text{REQUIREMENTS} + 0,1]$;
- $\text{min_cycle_length}_i \in [\text{MIN_CYCLE_LENGTH} - 6, \text{MIN_CYCLE_LENGTH} + 6]$;
- $\text{max_uptake}_i \in [\text{MAX_UPTAKE} - 0,25 ; \text{MAX_UPTAKE} + 0,25]$;
- $\text{max_outtake}_i \in [\text{MAX_OUTTAKE}]$;
- $\text{released_compounds}_i \in [\text{RELEASED_COUPOUNDS} - 0,00005 ; \text{RELEASED_COUPOUNDS} + 0,00005]$.

L'âge des cellules de la première génération est

- $\text{age}_i \in [0 ; \text{MIN_CYCLE_LENGTH}]$

V.6. CONCLUSION

L'élaboration du modèle et de la plateforme se sont nourris l'un l'autre. Au fur et à mesure que de nouveaux besoins se faisaient jour dans la construction du modèle de nouvelles fonctionnalités ont été implémentées dans la plateforme.

Le premier biais révélé concerne le type de balayage utilisé pour la mise à jour de l'automate cellulaire, ce qui nous a imposé de procéder à une mise à jour dans un ordre aléatoire. La reproductibilité du modèle a été vérifiée ainsi que son comportement sur différentes configurations matérielles et logicielles, mais la limite d'une cellule par maille était un frein considérable pour approfondir l'exploitation du modèle. Le comportement était bon, les trajectoires des simulations étaient bien celles attendues, mais l'effectif obtenu ne convenait pas, plafonnant à quelques dizaines de milliers d'individus au prix d'un temps de calcul important. Augmenter la taille de l'espace et le rayon de voisinage exploité faisait sens en théorie mais les ressources matérielles et durées de simulation devenaient insupportables. En effet une matrice de taille 500x500 avec un voisinage de Moore de rayon 1 autorise une simulation en une 15 de minutes sur une machine récente, ce temps est plus que triplé avec un voisinage de Moore de rayon 3, pour obtenir au maximum 250 000 cellules. Passer à un voisinage illimité nécessite 1 journée pour le même résultat, et en se contentant d'un voisinage de Moore de rayon 1 mais en doublant les dimensions de la matrice pour exploiter un espace de 1000x1000, une simulation dure alors plus d'une semaine avec un effectif final qui n'est au maximum que d'1 000 000 de cellules, soit 10^6 , on reste très loin des ordres de grandeur que l'on obtient *in vitro* avec des bactéries.

L'hypothèse du superindividu a alors fait sens et a permis d'outrepasser cette limite. Le rayon de voisinage n'a alors plus d'impact, et la taille de la matrice n'a plus beaucoup d'importance. Avec une matrice de 500x500 et un voisinage de Moore de rayon 1, les effectifs atteignent facilement 10^{10} individus dans un temps de calcul qui reste de l'ordre du quart d'heure.

Avec ce nouveau potentiel de simulation, les paramètres identifiés dans les études biologiques ont pu être exploités dans le modèle et ont permis d'obtenir des résultats en parfait accord avec les données biologiques.

On note toutefois la discontinuité qui s'observe dans les résultats de la simulation entre la phase exponentielle de croissance et la phase stationnaire, avec une chute brutale de la taille de la population. Dans le système expérimental *in vitro*, tous les individus sentiraient progressivement les ressources diminuer, et ralentiraient leur croissance en conséquence. La croissance aurait une valeur « continue » entre l'arrêt et le plein régime. Ceci est en partie lié au fait que la simulation est discrète, et que ce qui est donc évalué est le comportement moyen de la population.

La fluctuation importante qui apparaît lors du changement de régime dans la croissance observée lors de la simulation est liée à la discrétisation en termes d'espace et de temps. Le comportement observé est par paliers. Dans la phase stationnaire, on a une succession de phases de croissance en escalier et de chutes brutale de l'effectif. C'est déjà ce que nous avons observé avec Daisyworld, lors des changements de phases, la discrétisation rend le système instable avant qu'il n'oscille autour de son comportement moyen.

On peut aussi souligner que le modèle est simple. La phase stationnaire ne correspond pas à une règle explicite du modèle, pas plus que la croissance exponentielle. Les règles explicites disent qu'une cellule qui a pu dupliquer ses constituants et ne souffre pas peut se diviser, et qu'une cellule qui n'a pas suffisamment de ressources se délite petit à petit et non pas d'un coup. Ce comportement, phase exponentielle de croissance et phase stationnaire résulte d'un équilibre qui se forme par l'interaction des cellules et de leur environnement. Le paramétrage du modèle est très sensible, et la notion de recyclage du matériel cellulaire après la mort d'un individu n'a commencé à être introduite qu'après avoir vérifié et validé que la mort survenait au bon moment et comme attendu. L'identification des valeurs des paramètres est encore à affiner pour se rapprocher encore d'un comportement plus réaliste.

De plus, il n'y a pas de progressivité dans le changement de régime car l'ensemble des nutriments est accessible à chaque individu à chaque pas de temps de la simulation ; il y a des nutriments en assez grand nombre, un individu puise les derniers, et les individus mis à jour ensuite se trouvent en situation de carence. De fait, les individus croissent à plein régime, puis s'arrêtent, se dégradent et meurent. Le mode de croissance se trouve « discret » dans le sens où c'est du tout ou rien. Il s'agit d'une amélioration que nous envisageons d'expérimenter, par exemple en introduisant un second niveau de voisinage qui donnera aux cellules une visibilité des nutriments à portée, ou en « diluant » les nutriments dans l'espace afin de limiter ce qui est accessible à chacune des cellules à chaque instant, ou encore, en diminuant le pas de temps du modèle.

CHAPITRE VI. CONCLUSIONS ET PERSPECTIVES

L'idée de modéliser et simuler les effets des radiations sur les systèmes biologiques est un projet ambitieux. Les systèmes biologiques peuvent s'observer à plusieurs niveaux qui ne sont pas indépendants. Le fonctionnement d'une population et son développement dans le temps découle des interactions et relations entre les individus qui la constituent. Le bon fonctionnement d'un organisme et son adaptation à son milieu est lié au bon fonctionnement de ses organes. Même si l'on ne considère qu'un organisme unicellulaire, son bon fonctionnement est lié à la biochimie de son métabolisme sous le contrôle de son patrimoine génétique et sa relation à son environnement.

Pour modéliser l'effet des radiations sur les organismes, il faut pouvoir gérer des effets qui se passent à des échelles très différentes, en termes d'espace et en termes de temps. En termes d'espace, les phénomènes considérés se jouent à l'échelle atomique pour les interactions entre les rayonnements ionisants et les constituants cellulaires et à l'échelle de l'organisme pour la propagation des rayonnements. En termes de temps, les interactions physico-chimiques entre les rayonnements ionisants et les molécules cellulaires, et les réarrangements chimiques qui en découlent se déroulent sur des durées de l'ordre de la microseconde, alors que leur influence sur un organisme ou une population peut ne se faire sentir que plusieurs générations plus tard.

Traiter tous ces phénomènes nécessite un modèle de l'organisme capable de considérer aussi bien l'échelle moléculaire où se déroule la chimie du vivant, que l'échelle de la population où sont visibles les phénomènes d'adaptation et d'évolution. Or nous l'avons vu, ces phénomènes font l'objet de modèles différents, dont les approches sont difficilement compatibles.

Ces travaux s'inscrivent dans une dynamique à long terme et nous espérons avoir réussi à poser des bases qui permettront de répondre à la problématique de la modélisation en radiobiologie. En ayant fait le choix d'un modèle individu-centré, nous nous inscrivons dans une démarche plusieurs fois utilisée pour modéliser les systèmes biologiques, et qui a fait démonstration de son intérêt pour la modélisation et la simulation des systèmes complexes. Si pour l'instant notre choix s'est porté sur les automates cellulaires comme support à cette modélisation, nous nous accordons la possibilité de changer cette approche en développant une plateforme axée sur la simulation des modèles individu-centrés.

Même si nous ne considérons que la croissance de bactéries pour l'instant, le modèle proposé ici se veut prêt à accueillir les améliorations et précisions requises pour modéliser les processus se déroulant à échelle différentes. En effet, en optant pour trois niveaux de paramètres explicites, au niveau de la population, au niveau de l'individu, et à chaque instant, nous autorisons d'avoir des jeux de paramètres qui vont pouvoir varier sur ces trois échelles que sont l'espèce, l'organisme et ses possibilités de réponses à son environnement, alors que

ce dernier est souvent implicite dans les modèles de la littérature (BacSim, INDISIM) et ne permet pas de moduler ces trois niveaux indépendamment.

De plus, lors de la division cellulaire, nous prenons le parti de calculer les paramètres des cellules filles sur la base des paramètres de la cellule mère en priorité et non par rapport aux caractéristiques de l'espèce, ainsi si un changement dans une cellule survient, il pourra se propager dans la population avec la possibilité ainsi de simuler un processus évolutif.

L'échelle des processus biochimiques n'est actuellement représentée que par une fonction dont le résultat indique si oui ou non les choses se sont bien passées, autorisant ou non la cellule à poursuivre son développement. Le calcul de cette fonction doit être enrichi, voire couplé à un autre modèle. Le choix est encore à travailler, mais les modèles basés sur les réseaux de Petri comme Snoopy nous paraissent particulièrement adaptés. Les phases d'absorption et d'excrétion cellulaires correspondront alors à l'injection ou à la ponction de jetons dans le réseau de Petri. De plus, paramétrer dans un tel modèle chaque transition permettra de gérer les paramètres de chaque réaction individuellement, ce qui permettra dans le cadre d'enzymes disposant de plusieurs fonctions catalytiques d'impacter l'une indépendamment des autres en cas de mutation.

Ceci nous amène à considérer le développement du modèle sur les aspects génétiques. Pour l'instant, un simple jeu de paramètres représente le génome, comme BacSim ou INDISIM le font, mais il nous semble intéressant de tendre vers une gestion comme le permet Aevol et l'algorithme RBF-Gene. Le calcul du phénotype explicite correspondrait au paramétrage de chaque arc du réseau de Petri, et les trois opérateurs génétiques qu'ils emploient nous paraissent tout indiqués pour modéliser les mutations, qu'elles soient d'origine naturelle ou induites par un stress radiatif. Les aspects structurels du génome qui pourront être modifiés avec GEANT4-DNA trouveront alors à ce niveau leur utilisation au sein du modèle que nous proposons.

Arriver à développer le modèle proposé dans ces travaux pour étendre les modèles qui viennent d'être évoqués permettrait ainsi de rendre compte de toute la variabilité interindividuelle qui peut exister, en la basant sur des fondements génétiques. La variabilité pourrait en plus être modulée en permanence par la contrainte de l'environnement. La gestion actuelle de la population permettant une dérive des paramètres et autorisant un processus de création de variants comme on peut le trouver dans Avida par exemple, ce qui manque comme élément à déterminer, c'est de séparer les espèces. À quel moment de divergence génétique doit-on faire correspondre l'apparition d'une nouvelle espèce. De là se pose également la question de la reproduction sexuée pour laquelle il faut encore apporter des éléments de définition.

Parler de reproduction sexuée amène aussi la question d'organismes plus complexes que des bactéries, à commencer par des cellules eucaryotes. Si en termes de croissance et de gestion du métabolisme le parallèle sera facile à faire, la difficulté va porter sur la modélisation de la structure de la cellule avec ses nombreux organites. En effet, comment ses multiples surfaces membranaires se comportent elles avec les radiations, il reste des éléments à affiner. Une piste probablement intéressante sera de faire évoluer le support du modèle vers un modèle de Potts, permettant d'avoir de multiples unités de surface occupées par une seule et même cellule, avec la possibilité de gérer son volume, en poussant le fonctionnement pour que par

maille du CPM il n'y ait pas un, mais deux « tags », un pour la cellule à laquelle elle est rattaché, et un pour l'organite qui lui correspond.

Par extension, se posera la question du passage aux organismes pluricellulaires. Des organismes simples et bien connus, comme *Caenorhabditis elegans*, avec son nombre précis et relativement faible de cellules, 959, pourront encore être gérés simplement de la même manière qu'une population de cellules, les choses vont se compliquer pour des organismes plus complexes, et il sera temps d'y réfléchir, nous n'avons actuellement pas de solution à proposer pour passer cette étape. Peut-être faudra-t-il coupler d'autres types de modèles, ou compter sur des progrès techniques.

Nous pensons que le modèle proposé dans ces travaux constitue une base solide qui, avec les pistes d'évolution ici évoquées, permettra de modéliser les systèmes biologiques aux multiples échelles spatiales et temporelles nécessaires : de l'échelle moléculaire à l'échelle de l'organisme, en considérant des échelles de temps propres à correspondre aux processus chimiques à l'œuvre dans le métabolisme et les interactions avec l'environnement, et dans notre problématique l'action de la radioactivité, aux échelles de temps décrivant la vie de la population cellulaire avec le suivi des dérives génétiques qui peuvent apparaître, avec bien sûr à niveau intermédiaire, l'échelle de la cellule et de sa vie autonome.

En l'état, la définition du modèle permet de reproduire deux des grandes phases de la culture cellulaire : la phase exponentielle de croissance et la phase stationnaire, avec des résultats comparables aux résultats expérimentaux obtenus *in vitro*, aux résultats obtenus *in silico* avec d'autres modèles, et aux modèles à population. Sans changer le modèle, il doit être possible de simuler également la phase de latence, en changeant l'état des cellules de la population initiale. Des essais sont actuellement en cours pour vérifier cette assertion.

L'implémentation de la gestion de la diffusion des nutriments au sein de la plateforme devrait solutionner certains artefacts qui sont actuellement observés, en particulier les transitions entre les phases de la courbe de croissance.

BIBLIOGRAPHIE

- [1] G. Flesefeld et M. Groudine, «Controlling the double helix,» *Nature*, vol. 421, pp. 448-453, 2003.
- [2] L. Prescott, J. Harley, D. Klein, J. Willey, L. Sherwood et C. Woolverton, *Microbiologie*, DeBoeck, 2010.
- [3] R. Garret et C. Grisham, *Biochimie*, DeBoeck Université, 2000.
- [4] R. E. Buchanan, «Life phases in a bacterial culture,» *The Journal of Infectious Diseases*, vol. 23(2), pp. 109-125, 1918.
- [5] A. Lesne, «Biologie des systèmes : l'organisation multiéchelle des systèmes vivants,» *Médecine sciences*, vol. 25, pp. 585-587, 2009.
- [6] G. Beslon, «Apprivoiser la vie : Modélisation individu-centrée de systèmes biologiques complexes,» 2008.
- [7] «Stumble Upon via BioNumbers,» 27 Mai 2012. [En ligne]. Available: http://www.stumbleupon.com/su/2HBT3v/www.random-science-tools.com/chemistry/chemical_comp_of_body.htm#.
- [8] «BioNumbers,» [En ligne]. Available: www.bionumbers.org.
- [9] H. Bossel, *Modeling and Simulation*, Springer, 1994.
- [10] IAEA, *Model for Analysis of Energy Demand*, 2006.
- [11] R. C. Whiting et R. E. Buchanan, «Letter to the Editor : A classification of models in predictive microbiology. A reply to K.R. Davey,» *Food Microbiology*, vol. 10, pp. 175-177, 1993.
- [12] T. Ross et P. Dalgaard, *Modeling microbial responses in food : secondary models*, CRC Press, 2003.
- [13] T. Malthus, «Essay on principle of population,» 1798.
- [14] T. Malthus, *On Population*, 1809.
- [15] R. L. Buchanan, R. C. Whiting et W. C. Damert, «When is simple good enough : a comparison of the Gompertz, Baranyi and Three-phases lienar models for fitting bacterial growth curves,» *Food microbiology*, vol. 14, pp. 313-326, 1997.

- [16] P.-F. Verhulst, «Notice sur la loi que poursuit la population dans son accroissement,» *Correspondance Mathématique et Physique*, vol. 10(4), pp. 113-121, 1838.
- [17] B. Gompertz, «On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies,» *Philosophical Transactions of the Royal Society of London*, vol. 115, pp. 513-585, 1825.
- [18] M. H. Zwietering, I. Jongenbruger, F. Rombouts et K. Van'T Riet, «Modeling bacterial growth curve,» *Applied and Environmental Microbiology*, vol. 56(6), pp. 1875-1881, 1990.
- [19] J. Baranyi, P. J. McClure, J. P. Sutherland et T. A. Roberts, «Modeling bacterial growth responses,» *International Journal of Microbiology*, vol. 12, pp. 190-194, 1993.
- [20] L. Delhalle, G. Daube, I. Adolphe, S. Creveceour et A. Clinquart, «Les modèles de croissance en microbiologie prévisionnelles pour la maitrise et la sécurité des aliments (synthèse bibliographique),» *BASE Biotechnology, Agronomy, Society and Environment*, vol. 16(3), pp. 369-381, 2012.
- [21] M. Scheffer, J. M. Baveco, D. L. De Angelis, K. A. Rose et E. H. van Nes, «Super-individuals a simple solution for modeling large populations on an individual basis,» *Ecological Modeling*, vol. 80, pp. 161-170, 1995.
- [22] B. Ermentrout et L. Edelstein-Keshet, «Cellular automata approaches to biological modeling,» *Journal of Theoretical Biology*, vol. 160, pp. 97-133, 1993.
- [23] M. Alber, M. Kiskowski, J. Glazier et Y. Jiang, «On cellular automaton approaches to modeling biological cells,» *Mathematical Systems Theory in Biology, Communications, Computation, and Finance*, vol. 134, pp. 1-39, 2003.
- [24] A. Deutsch et S. Dormann, *Cellular Automaton Modeling of Biological pattern formation*, Birkhäuser, Éd., Birkhäuser, 2004.
- [25] J. Von Neuman et A. Burks, *Theory of self replicatng cellular automata*, University of Illinois Press, 1966.
- [26] M. Gardner, «The fantastic combinations of John Conway's new solitaire game "life",» *Scientific American*, vol. 223, pp. 120-123, Octobre 1970.
- [27] J. Albert et K. Culik, «A Simple Cellular Automaton and its One-Way and Totalistic Version,» *Complex Systems*, vol. 1, pp. 1-16, 1987.
- [28] M. Cook, «Universality in Elementary Cellular Automata,» *Complex Systems*, vol. 15, pp. 1-40, 2004.
- [29] S. Dormann et A. Deutsch, «Modeling of Self-Organized Avascular Tumor Growth with a Hybrid Cellular Automaton,» *In Silico Biology*, vol. 2(3), pp. 393-406, 2002.
- [30] S. Maheswaran, P. Hammond et P. M. Speight, «Modeling Epithelial Cell Behavior

- Organization,» *NanoBioscience*, vol. 6(1), pp. 77-85, 2007.
- [31] B. Laforge, D. Guez, M. Martinez et J. J. Kupiec, «Modeling embryogenesis and cancer : an approach based on an equilibrium between the autostabilization of stachastic gene expression and the interdependance,» *Progress in Biophysics and Molecular Biology*, vol. 89, pp. 93-120, 2005.
- [32] M. Markus, D. Böhm et M. Schmick, «Simulation of vessel morphogenesis using,» *Mathematical Biosciences*, vol. 156, pp. 191-206, 1999.
- [33] G. Cocho, R. Pérez-Pascual, J. Rius et F. Soto, «Discrete Systems, Cell-Cell Interactions and Color Pattern of Animals - Clonal Theory and Cellular Automata,» *Journal of Theoretical Biology*, vol. 125, pp. 437-447, 1987.
- [34] I. Kusch et M. Markus, «Mollusc shell pigmentation : Cellular automaton simulations and evidence for undecidability,» *Journal of Theoretical Biology*, vol. 178, pp. 333-340, 1995.
- [35] I. Karakyllidis et A. Thanailakis, «A model for predicting forest fire spreading using cellular automata,» *Ecological Modeling*, vol. 99, pp. 87-97, 1997.
- [36] A. Alexandridis, D. Vakalis, C. Siettos et G. Bafas, «A cellular automata model for forest fire spread prediction : The case of the wildfire that swept through Spetses Island in 1990,» *Applied Mathematics and Computation*, vol. 204, pp. 191-201, 2008.
- [37] S. H. White, Á. M. Del Rey et G. R. Sánchez, «Modeling epidemics using cellular automata,» *Applied Mathematics and Computation*, vol. 186, pp. 193-202, 2007.
- [38] S. Ching Fu, «Modeling epidemic spread using Cellular Automata,» University of Western Australia, 2002.
- [39] H. Meinhardt, «Max Planck Intute for Developmental Biology - Hans Meinhardt,» [En ligne]. Available: <http://www.eb.tuebingen.mpg.de/research/emeriti/hans-meinhardt/shells.html>.
- [40] R. B. Potts, «Some generalized order-disorder transformations,» *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48(1), pp. 106-109, 1952.
- [41] J. Glazier et F. Graner, «Simulation of the differential adhesion driven rearrangement of biological cells,» *Physical Review E*, vol. 47(3), pp. 2128-2154, 1993.
- [42] F. Graner et J. Glazier, «Simulation of Biological Cell Sorting Using a Two-Dimensional Extended Potts Model,» *Physica Review Letters*, vol. 69(13), pp. 2013-2017, 1992.
- [43] Y. Jiang, H. Levine et J. Glazier, «Possible cooperation of differential adhesion and chemotaxis in mound formation of *Dicyostelium*,» *Biophysical Journal*, vol. 75, pp. 2615-2625, 1998.
- [44] N. Savill et P. Hogeweg, «Modelling Morphogenesis] From Single Cells to Crawling

- Slugs,» *Journal of Theoretical Biology*, vol. 184, pp. 229-235, 1997.
- [45] J. Hardy, O. de Pazzis et Y. Pomeau, «Molecular dynamics of a classical lattice gas : Transport properties and time correlation functions,» *Physical Review A*, vol. 13(5), pp. 1949-1961, 1976.
- [46] U. Frisch, B. Hasslacher et Y. Pomeau, «Lattice-Gas Automata for the Navier-Stokes equation,» *Physical Review Letters*, vol. 56(14), pp. 1505-1508, 1986.
- [47] D. A. Wolf-Gladrow, *Lattice-Gas Cellular Automata and Lattice Boltzmann Models - An introduction*, Springer, 2005.
- [48] S. Wolfram, «Cellular automaton fluids - basic theory,» *Journal of Statistical Physics*, Vols. 45 sur 91(3-4), pp. 471-526, 1986.
- [49] S. Wolfram, «Theory and Applications of Cellular Automata,» *Advanced Series on Complex Systems*, vol. 1, 1986.
- [50] R. Ouared, B. Chopard, B. Stahl, D. A. Rüfenacht, H. Yilmaz et G. Courbebaisse, «Thrombosis modeling in intracranial aneurysms : a lattice Boltzmann numerical algorithm,» *Computer Physics Communications*, vol. 179, pp. 128-131, 2005.
- [51] B. Chopard, R. Ouared, D. A. Rüfenacht et H. Yilmaz, «Lattice Boltzmann modeling of thrombosis in giant aneurysm,» *International Journal of Modern Physics C*, vol. 18(4), pp. 712-721, 2007.
- [52] H. Hatzikirou, L. Brusch, C. Schaller, M. Simon et A. Deutsch, «Prediction of traveling front behavior in a lattice-gas cellular automaton model for tumor invasion,» *Computers and Mathematics with Applications*, vol. 59, pp. 2326-2339, 2010.
- [53] V. Grimm, «Ten years of individual-based modeling in ecology : what have we learned and what could we learn in the future,» *Ecological Modeling*, vol. 115, pp. 129-148, 1999.
- [54] F. Amblard et D. Phan, *Modélisation et simulation multi-agents : Applications pour les Sciences de l'Homme et de la Société*, Hermes Science Publications, 2006.
- [55] F. Amigoni et V. Schiaffonati, «Multiagent-based simulations in Biology,» *Model-based Reasoning in Science, Technology and Medicine : Studies in Computational Intelligence*, vol. 64, pp. 179-191, 2007.
- [56] V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jorgensen, W. M. Mooij, B. Müller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R. A. Stillman, R. Vabo, U. Visser et D. L. De Angelis, «A standard protocol for describing individual-based and agent-based models,» *Ecological Modeling*, vol. 198, pp. 115-226, 2006.
- [57] V. Ginot et C. Le Page, «Mobidyc, a generic multi-agents simulator for modeling

- populations dynamics,» *Lecture Notes in Computer Science*, vol. 1416, pp. 805-814, 1998.
- [58] T. Brochier, J.-M. Ecoutin, L. T. De Morais, D. M. Kaplan et R. Lae, «A multi-agent ecosystem model for studying changes in a tropical estuarine fish assemblage within a marine protected area,» *Aquatic Living Resources*, vol. 26(2), pp. 147-158, 2012.
- [59] S.-J. Lin et P.-F. Chang, «Using swarm to build a multi-agent simulation system for artificial stock market,» *Proceeding of the second Workshop on Knowledge Economy and Electronic Commerce*, vol. 1, 2004.
- [60] S. Ilie et C. Badica, «Mutli-agent based distributed framework for swarm intelligence,» *Procedia Computer Science*, vol. 18, pp. 611-620, 2013.
- [61] L. Bortolussi, A. Dovier et F. Fogolari, «Multi-Agent Simulation of Protein Folding,» *AMAS Workshop on multi-Agent Systems for Medecine, Computational Biology, and Bioinformatics*, vol. 1, pp. 91-105, 2005.
- [62] A. Orro, M. Saba, E. Vargiu et G. Mancosu, «Using a personalized, adaptive and cooperative multiagent system to predict protein secondary structure,» *AAMAS Workshop on multi-Agent Systems for Medecine, Computational Biology, and Bioinformatics*, vol. 1, pp. 170-183, 2005.
- [63] S. Bandini, S. Manzoni et V. Giuseppe, «Immune system modeling with situated Cellular Agents,» *AMAS Workshop on multi-Agent Systems for Medecine, Computational Biology, and Bioinformatics*, vol. 1, pp. 78-90, 2005.
- [64] T. Bosse, C. M. Jonker et J. Treur, «Modeling the dynamics of intracellular process as an organisation of multiple agents,» *AMAS Workshop on multi-Agent Systems for Medecine, Computational Biology, and Bioinformatics*, vol. 1, pp. 107-121, 2005.
- [65] F. Bousquet, C. Le page et J.-P. Müller, «Modélisation et simulation multi-agent,» *Actes des deuxièmes assises nationales du GdR I3*, 2002.
- [66] M. Hwang, M. Garbey, S. Berceci et R. Tran-Son-Tay, «Rule-based simulation of multicellular biological systems - a review of modeling techniques,» *Cellular and Molecular Bioengineering*, vol. 2(3), pp. 285-294, 2009.
- [67] «SimCell,» [En ligne]. Available: <http://wishart.biology.ualberta.ca/SimCell/>.
- [68] D. S. Wishart, R. Yang, D. Arndt, P. Tang et J. Cruz, «Dynamic cellular automata : an alternative approach to cellular simulation,» *In Silico Biology*, vol. 4(15), 2004.
- [69] «V-Cell-1,» [En ligne]. Available: <http://vcell.org/>.
- [70] «V-Cell-2,» [En ligne]. Available: <http://www.nrcam.uchc.edu/>.
- [71] J. Schaff, C. C. Fink, B. Slepchenko, J. H. Carson et L. M. Loew, «A general computational framework for modeling cellular structure and function,» *Biophysical Journal*, vol. 73,

- pp. 1135-1146, 1997.
- [72] «V-Cell_tuto,» [En ligne]. Available: <http://vcell.org/index.html>.
- [73] B. M. Slepchenko, I. Semenova, I. Zaliapin et V. Rodionov, «Switching of membrane organelles between cytoskeletal transport systems is determined by regulation of the microtubule-based transport,» *Journal of Cellular Biology*, vol. 179(4), pp. 635-641, 2007.
- [74] J. A. Ditlev, N. M. Vacanti, I. L. Novak et L. M. Loew, «An Open Model of Actin Dendritic Nucleation,» *Biophysical Journal*, vol. 96(9), pp. 3529-3542, 2009.
- [75] S. A. Brown, I. I. Moraru, J. C. Schaff et L. M. Loew, «Virtual NEURON: a strategy for merged biochemical and electrophysiological modeling,» *Journal of Computational Neuroscience*, pp. 385-400, 2011.
- [76] computer science department, «data structures and software dependability,» brandenburg university of technology cottbus, [En ligne]. Available: <http://www-dssz.informatik.tu-cottbus.de/DSSZ/Software/Software>.
- [77] C. A. Petri, «Kommunikation mit automaten,» 1962.
- [78] M. Heiner, R. Richter, M. Schwarick et C. Rohr, «Snoopy - a tool to design and animate/simulate graph-based formalisms,» *SIMUTools 2008*, 2008.
- [79] W. Marwan, C. Rohr et M. Heiner, *Methods in Molecular Biology*, Humana Press, 2012.
- [80] J.-U. Kreft, G. Booth et J. W. T. Wimprenny, «BacSim, a simulator for individual-based modelling of bacterial colony growth,» *Microbiology*, vol. 144, pp. 3275-3287, 1998.
- [81] G. Booth, «Gecko : a continuous 2-D world for ecological modeling,» *Artificial Life*, vol. 3, pp. 147-163, 1997.
- [82] J.-U. Kreft, «BacSim,» [En ligne]. Available: http://www.theobio.uni-bonn.de/people/jan_kreft/bacsim.html.
- [83] E. J. Dens, K. Benaerts, A. R. Standaert et J. F. Van Impe, «Cell division theory and individual-based modeling of microbial lag : Part I. The theory of cell division,» *International Journal of Food Microbiology*, vol. 101, pp. 303-318, 2005.
- [84] E. J. Dens, K. Benaerts, A. R. Standaert, J. U. Kreft et J. F. Van Impe, «Cell division theory and individual-based modeling of microbial lag : Part II. Modeling lag phenomena induced by temperature shifts,» *International Journal of Food Microbiology*, vol. 101, pp. 319-332, 2005.
- [85] M. Ginovart, D. Lopez et J. Valls, «INDISIM, An Individual-based Discrete simulations Model to study bacterial cultures,» *Journal of Theoretical Biology*, vol. 214, pp. 305-319, 2002.

- [86] C. Prats, D. Lopez, A. Giro, J. Ferrer et J. Valls, «Individual-based modeling of bacterial cultures to study the microscopic causes of the lag phase,» *Journal of Theoretical Biology*, vol. 241, pp. 939-953, 2006.
- [87] M. Ginvar, D. Lopez et A. Gras, «Individual-based modeling of microbial activity to study mineralization of C and N nitrification process in soil,» *Nonlinear Analysis : Real World Applications*, vol. 6, pp. 773-795, 2005.
- [88] A. Gras et M. Ginovart, «INDISIM-SOM, an individual-based model to study short-term evolutions of carbon and nitrogen pools-related to microbial activity in soil organic matter,» *Proceedings 20th European Conference on Modeling and Simulation*, 2006.
- [89] M. Ginovart, J. Xifré, D. Lopez et M. Silbert, «INDISIM-YEAST, an individual-based model to study yeast population in batch cultures,» *Communicating Current Research and Educational Topics and Trends in Applied Microbiology*, pp. 401-409, 2007.
- [90] X. Portell, A. Gras et M. Ginovart, «INDISIM-Saccha, an individual-based model to tackle *Saccharomyces cerevisiae* fermentations,» *Ecological Modeling*, vol. 279, pp. 12-23, 2014.
- [91] «Terra Home Page,» [En ligne]. Available: <http://life.ou.edu/tierra/>.
- [92] T. S. Ray, «Evolution and optimization of digital organisms,» *Scientific Excellence in Supercomputing*, pp. 486-531.
- [93] K. Sims, «Evolving Virtual Creatures,» *Computer Graphics*, pp. 15-22, 1994.
- [94] K. Sims, «Evolving 3D morphology and behavior by competition,» *Artificial Life Proceedings*, vol. 4, pp. 28-39, 1994.
- [95] «Virtual Life Software,» [En ligne]. Available: <http://life.ou.edu/VirtualLife/>.
- [96] T. Ray, «Aesthetically Evolved Virtual Pets,» *Leonardo, Journal of the International Society for the Arts, Sciences and Technology*, vol. 34(4), pp. 313-316, 2001.
- [97] C. Adami et C. T. Brown, «Evolutionary learning in 2D artificial life system "Avida",» *Proceedings of Artificial Life IV*, vol. 1194, pp. 377-381, 1994.
- [98] «Avida,» [En ligne]. Available: <http://avida.devosoft.org/>.
- [99] R. Lenski, C. Ofria, R. Pennock et C. Adami, «The evolutionary origin of complex features,» *Nature*, vol. 423, pp. 139-144, 2003.
- [100] R. Pennock, «Darwin's place on campus is secure - but not supreme,» *Science*, vol. 311(5762), pp. 769-771, 2006.
- [101] R. Pennock, «Learning evolution an the nature of science using evolutionary computing and artificial life,» *McGill Journal of Education*, vol. 42(2), pp. 211-224, 2007.

- [102] R. Pennock, «AvidaED,» [En ligne]. Available: <http://avida-ed.msu.edu/>.
- [103] «Aevol,» [En ligne]. Available: <http://www.aevol.fr/>.
- [104] G. Beslon, C. Knibbe, H. Soula et J.-M. Fayard, «The RBF-Gene Model,» *Artificial Neural Nets and Genetic Algorithms*, pp. 187-192, 2003.
- [105] V. Lefort, C. Knibbe, G. Beslon et J. Favrel, «A bio-inspired genetic algorithm with a self-organizing genome : The RBF-Gene Model,» *Lecture Notes in Computer Sciences*, pp. 406-407, 2004.
- [106] C. Knibbe, o. Mazet, F. Chaudier, J.-M. Fayard et G. Beslon, «Evolutionary coupling between the deleteriousness of gene mutations and the amount of non-coding sequences,» *Journal of Theoretical biology*, vol. 244, pp. 621-630, 2006.
- [107] B. Batut, D. Parsons, S. Fischer, G. Beslon et C. Knibbe, «In-silico experimental evolution : a tool to test evolutionary scenarios,» *Proceeding of the Eleventh annual Research in Computational Molecular Biology*, 2013.
- [108] H. Kitano, foundations of Systems Biology, MIT Press, 2001.
- [109] M. Hucka, A. Finney, H. Sauro et H. Bolouri, «Systems Biology Markup Language (SBML) Level 1 : Structures and facilities for basic model definitions,» 2001.
- [110] M. Hucka, A. Finney, H. Sauro, H. Bolouri, J. Doyle, H. Kitano, A. Arkin, B. Bornstein, D. Bray, A. Cornish-Bowden, A. Cuellar, S. Dronov, E. Gilles, M. Ginkel, V. Gor, I. Goryanin, W. Hedley, T. Hodgman, J. H. Hofmeyr, P. Hunter, N. Juty, J. Kasberger, A. Kremling, U. Kummer, N. Le Novère, . L. Loew, D. Lucio, P. Mendes, E. Minch, E. Mjolsness, Y. Nakayama, M. Nelson, P. Nielsen, T. Sakurada, J. Schaff, B. Shapiro, T. Shimizu, H. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner et J. Wang, «The systems biology markup language (SBML) : a medium for representation and exchange of biochemical network models,» *Bioinformatics*, vol. 19(4), pp. 524-531, 2003.
- [111] M. Hucka, F. Bergmann, S. Hoops, S. Keating, S. Sahle, J. Schaff, L. Smith et D. Wilkinson, «The Systems Biology Markup Language (SBML) : Language Specification for Level 3 Version 1 Core,» 2010.
- [112] «CellDesigner,» [En ligne]. Available: <http://www.celldesigner.org/>.
- [113] Bioengineering Institute, Auckland University, «CellML,» [En ligne]. Available: <https://www.cellml.org/>.
- [114] W. Hedley, P. Nielsen et P. Hunter, «XML Languages for describing biological models and data,» *Proceedings of Biomedical Engineering Society Annual Meeting*, 2000.
- [115] «OpenCell,» [En ligne]. Available: <http://www.cellml.org/tools/opencell/>.
- [116] Department of Computational and Systems Biology, «M-Cell,» [En ligne]. Available: <http://mcell.org/index.html>.

- [117] «E-Cell,» [En ligne]. Available: <http://www.e-cell.org/>.
- [118] K. Yugi et M. Tomita, «A general computational model of mitochondrial metabolism in a whole organelle scale,» *Bioinformatics Applications Note*, vol. 11(20), pp. 1795-1796, 2004.
- [119] A. Yachie-Kinoshita, T. Nishino, H. Shimo, M. Suematsu et M. Tomita, «A metabolic model of human erythrocytes : practical application of the E-Cell simulation environment,» *Journal of Biomedecine and Biotechnology*, 2010.
- [120] P. Smolen, P. Hardin, B. Lo, D. Baxter et J. Byrne, «Simulation of Drosophila circadian oscillations, mutations, and light responses by a model with VRI, PDP-1, and CLK,» *Biophysical Journal*, vol. 86, pp. 2786-2802, 2004.
- [121] J. A. Izaguirre, R. Chaturvedi, C. Huang, T. Cickowski, J. Coffland, G. Thomas, G. Forgacs, M. Alber, G. Hentschel, S. Newman et J. A. Glazier, «CompuCell, a multi-model framework for simulation of morphogenesis,» *Bioinformatics*, vol. 20(7), pp. 1129-1137, 2004.
- [122] «CompuCell3D,» [En ligne]. Available: <http://www.compuCell3d.org/>.
- [123] N. Poplawski, A. Shirinifard, M. Swat et J. Glazier, «Simulation of single-species bacterial-biofilm growth using the Glazier-Grane-Hogeweg model and the CompuCell3D modeling environment,» *Mathematical Biosciences and Engineering*, vol. 5(2), pp. 355-388, 2008.
- [124] B. Vasiev, A. Balter, M. Champlain, J. Glazier et C. Weijer, «Modeling gastrulation in the chick embryo : formation of the primitive streak,» *PLOS One*, vol. 5(5), 2010.
- [125] S. Hester, J. Belmonte, S. Gens, S. Clendenon et J. Glazier, «A multi-cell, multi-scale model of vertebrate segmentation and somite formation,» *PLOS Computational Biology*, vol. 7(10), 2011.
- [126] A. Dias, I. de Almeida, J. Belmonte, J. Glazier et C. Stern, «Somites without a clock,» *Science*, vol. 343(6172), pp. 791-795, 2014.
- [127] M. Swat, G. Thomas, J. Belmonte, A. Shirinifard, D. Hmeljak et J. Glazier, «Multi-scale modeling of tissues using CompuCell3D,» *Methods in Cellular Biology*, vol. 110, pp. 325-366, 2012.
- [128] «Morphex,» [En ligne]. Available: <http://morphex.org/tiki-index.php?page=Home>.
- [129] N. Britton, *Essential Mathematical Biology*, Springer, 2003.
- [130] M. Gardner, «The fantastic combinations of John Conway's new solitaire game "life",» *Scientific American*, vol. 223, pp. 120-123, Octobre 1970.
- [131] A. K. Dewdney, «Computer recreations: The cellular automata programs that create Wireworld, Rugworld and other diversions,» *Scientific American*, vol. 262, pp. 146-149,

1990.

- [132] «The Wireworld Computer,» [En ligne]. Available: <http://www.quinapalus.com/wi-index.html>.
- [133] «Astrosurf,» [En ligne]. Available: <http://www.astrosurf.com/luxorion/theorie-gaia2.htm>.
- [134] F. Aiboud, «Méthodes approchées pour la résolution de problèmes inverses : Identification paramétrique et génération de formes,» 2013.
- [135] M. Scheffer, J. M. Baveco, D. L. De Angelis, K. A. Rose et E. H. van Nes, «Super-individuals a simple solution for modeling large populations on an individual basis,» *Ecological Modeling*, vol. 80, pp. 161-170, 1995.
- [136] C. Nicolas, «Analyse globale de la réponse adaptative du métabolisme central d'E.coli à une modulation contrôlée de la glucose-6-phosphate deshydrogénase : étude par fluxomique et corrélation transcriptome / fluxome,» Thèse de Doctorat de l'Institut National des Sciences Appliquées de Toulouse, 2007.
- [137] «BioNumbers,» [En ligne]. Available: www.bionumbers.org. [Accès le décembre 2012].
- [138] R. Phillips, J. Kondev et J. Theriot, *Physical Biology of the Cell*, Garland Science, 2008.

GLOSSAIRE

Acidophile	Caractéristique d'un organisme dont le développement est le plus efficace à ph acide (<7).
Adn	Acide désoxyribonucléique, c'est la molécule support de l'information génétique.
Alcalophile	Caractéristique d'un organisme dont le développement est le plus efficace à ph alcalin (>7).
Allèle	Nom désignant une version d'un gène. Lorsque un individu portent plusieurs copies d'un gène, soit il possède deux fois le même allèle, il est dit homozygote pour cet allèle, soit il porte deux allèles différents, il est alors dit hétérozygote pour ce gène.
Anabolisme	Ensemble des réactions biochimiques de fabrication se déroulant au sein d'une cellule.
Antiport	Cotransport dans lequel les deux composés sont transportés dans des sens opposés.
Apoptose	Mort programmée, c'est un mécanisme de suicide cellulaire.
Appareil De Golgi	Organite rencontré chez toutes les cellules eucaryotes, son rôle s'apparente à celui d'une gare de triage. C'est lui qui va permettre d'envoyer les protéines produites, après leur maturation dans le réticulum endoplasmique, à l'endroit où elles sont requises au sein de la cellule.
Aquaporine	Porine spécifique de la diffusion de l'eau. Note : pour la diffusion de l'eau, on préfère parler d'osmose.
Archaeobactérie	Règne du vivant regroupant des cellules procaryotes capables de survivre dans des milieux très hostiles pour tout autre organisme vivant.
Arn	Acide ribonucléique, son rôle est principalement l'échange d'information entre la molécule d'ADN et les ribosomes.
Bioréacteur	Système dans lequel le milieu de culture est renouvelé en continu. Du milieu frais est ajouté en continu à la culture pendant que du milieu usé est ponctionné au même rythme, ceci afin de maintenir l'organisme cultivé dans un milieu favorable et maintenir le plus longtemps possible une croissance efficace.
Catabolisme	Ensemble des réactions biochimiques de dégradation se déroulant au sein d'une cellule.
Centrosome	Organite central dans l'organisation des protéines de structure (cytosquelette) au sein d'une cellule animale.
Chémostat	Voir bioréacteur.

Chloroplaste	Organite spécifique du règne végétal; il est spécialisé dans la réalisation de la photosynthèse.
Chromatine	Forme normale sous laquelle se présente le matériel génétique dans une cellule. Elle est constituée d'une succession de nucléosomes. Lorsque la chromatine est peu enroulée, on parle d'euchromatine, lorsqu'elle est super enroulée, on parle d'hétérochromatine.
Chromosome	Structure constituée de chromatine.
Commensalisme	Mode d'alimentation d'un organisme qui se nourrit de débris de repas ou de parasite d'un organisme hôte plus grand, sans lui faire de tort, voire même en lui rendant un service.
Cotransport	Mécanisme de transport au cours duquel l'énergie générée par le transport d'un élément suivant son gradient électrochimique est mise à profit pour le transport d'un autre élément. Le transport des deux composés peut se faire dans le même sens, voir symport, ou dans des sens opposés, voir antiport.
Culture En Batch	Culture en milieu non renouvelé. Le milieu est introduit et inoculé en début de culture, il n'est pas modifié autrement que par les organismes cultivés en son sein jusqu'à l'arrêt de la culture.
Cytoplasme	Désigne tout le contenu de la membrane plasmique, c'est-à-dire, tout le contenu d'une cellule. Il diffère du cytosol dans le sens que le cytosol ne désigne que la solution saline, alors que cytoplasme désigne également les organites, organelles, grains, etc...
Cytosol	Désigne la solution saline contenue par la membrane plasmique, c'est-à-dire la fraction liquide et les éléments qui sont dissous dedans, d'une cellule. Il diffère du cytoplasme dans le fait que le cytoplasme désigne également les éléments insolubles.
Densité Optique	Mesure de la quantité de lumière absorbée par une solution. Très simple à mettre en œuvre; et très utilisée dans les études in vitro, cette unité de mesure repose sur la loi de Beer Lambert.
Drépanocytose	Egalement appelée anémie falciforme, cette maladie est liée à une mutation du gène codant l'hémoglobine qui se traduit par une protéine qui polymérise et forme des fibres. Les globules rouges sont alors déformés prenant la forme de faucille donne son nom à la maladie. Les globules ainsi déformés perdent leur élasticité, il en résulte qu'ils ont tendance à obstruer les capillaires sanguins, gênant la circulation et provoquant des thromboses, et à se rompre. Cette dernière caractéristique est responsable de l'hémolyse qui mène à l'anémie.
Eubactérie	Règne du vivant regroupant toutes les cellules que l'on appelle couramment bactéries.
Eucaryote	Désigne les cellules dotées d'un noyau délimité par une double membrane.
Euchromatine	Forme faiblement condensée de la chromatine qui autorise l'accès aux séquences génétiques pour leur expression dans le cadre du fonctionnement cellulaire.

Eumycète	Nom des organismes du règne fongique que l'on appelle couramment "champignons". Ce règne ne se limite regroupe également les champignons unicellulaires (levures) et filamenteux (moisissures).
Gamète	Cellule spécifique rencontrée uniquement chez les organismes ayant recours à la reproduction sexuée, on les appelle aussi cellules sexuelles ou germinales. Elles ont la particularité de ne porter qu'un demi-patrimoine génétique. La fécondation est nécessaire pour mettre en présence deux gamètes, mâle et femelle, portant chacun un demi-génome, pour former un nouvel individu doté d'un génome complet.
Génome	Ensemble du matériel génétique d'un individu, c'est-à-dire l'ensemble de ses séquences génétiques. Couramment, il s'agira de désigner l'ensemble de ses chromosomes. Remarque : chez les organismes qui ne possèdent qu'un chromosome unique, on parle indifféremment de chromosome ou de génome.
Génomique	Etude du fonctionnement d'un système vivant par analyse de l'ensemble de son génome.
Génotype	Composition génétique de l'individu, c'est-à-dire sa composition allélique (voir allèle).
Glycocalyx	Couche de sucres produites par les cellules eucaryotes autour de leur enveloppe pour "créer" ou maintenir un milieu qui leur est favorable, pour faciliter leur adhésion à leur substrat, pour se protéger, ou pour faciliter des échanges avec l'environnement.
Halophile	Caractéristique d'un organisme qui réclame pour son développement optimal, une forte concentration en sels dans son milieu.
Halotolérant	Caractéristique d'un organisme qui parvient à se développer dans un milieu dont la concentration en sels est élevée. Un tel organisme se développe toutefois moins vite dans de telles conditions qu'en présence d'un milieu dont la concentration en sels sera moindre.
Haptotaxie	"Déplacement par toucher"; ce terme désigne un mode de déplacement des cellules par déformation de leur enveloppe qui mène à la formation de pieds. Ces pieds sont "étendus", puis posés, et la cellule les utilise alors pour se tracter.
Hétérochromatine	Forme fortement condensée de la chromatine qui interdit l'accès aux séquences génétique par les enzymes permettant leur expression. Les séquences dans l'hétérochromatine sont "éteintes". Cette forme peut être constitutive et permet de ne pas exprimer des gènes inutiles à la cellule spécialisée (par exemple un neurone n'exprime pas les gènes nécessaires au fonctionnement d'une cellule du foie).
Hétérozygote	Caractéristique d'un individu qui présente une hétérozygotie pour un gène, c'est-à-dire qui porte deux allèles différents pour un même gène.
Hétérozygotie	Etat résultant de la présence de deux allèles différents pour un gène.
Histone	Protéine assurant la structuration tridimensionnelle de la fibre d'ADN. La double hélice constituée de deux brins d'ADN s'enroule autour des histones comme un fil autour d'une bobine pour former un nucléosome.

Homozygote	Caractéristique d'un individu qui présente une homozygotie pour un gène, c'est-à-dire qui porte deux copies d'un même allèle.
Homozygotie	Etat résultant de la présence de deux allèles identiques pour un gène.
Hyperthermophile	Caractéristique d'un organisme pour lequel les températures moyennes (environ 80-110 °C) sont les plus favorables au développement.
Hypertonique	Se dit d'un milieu qui, du fait de sa forte concentration en sels ou autres composés, exerce une forte pression osmotique sur l'organisme étudié. Cela se traduit par un mouvement d'eau de l'organisme vers le milieu.
Hypotonique	Se dit d'un milieu qui, du fait de sa faible concentration en sels ou autres composés, exerce une faible pression osmotique sur l'organisme étudié. Cela se traduit par un mouvement d'eau du milieu vers l'organisme étudié.
<i>In Silico</i>	Littéralement "dans le silicium", se dit d'un essai réalisé à l'aide de l'ordinateur.
<i>In Vitro</i>	Littéralement "dans le verre", se dit d'un essai réalisé en laboratoire dans un tube à essai à partir d'un élément isolé d'un organisme. <i>Remarque</i> : parfois, la distinction entre <i>in vitro</i> et <i>in vivo</i> est légère, par exemple, étudier une enzyme extraite d'une cellule animale dans un tube à essai est une expérience <i>in vitro</i> , étudier cette même enzyme dans le corps de l'animal est une expérience <i>in vivo</i> , mais l'étude de cette même enzyme dans une culture d'une lignée de cellule de l'animal pourra tantôt être considérée comme étant <i>in vitro</i> ou <i>in vivo</i> .
<i>In Vivo</i>	Littéralement "dans le vivant", se dit d'un essai réalisé au sein d'un être vivant. <i>Remarque</i> : parfois, la distinction entre <i>in vitro</i> et <i>in vivo</i> est légère, par exemple, étudier une enzyme extraite d'une cellule animale dans un tube à essai est une expérience <i>in vitro</i> , étudier cette même enzyme dans le corps de l'animal est une expérience <i>in vivo</i> , mais l'étude de cette même enzyme dans une culture d'une lignée de cellule de l'animal pourra tantôt être considérée comme étant <i>in vitro</i> ou <i>in vivo</i> .
Isotonique	Se dit d'un milieu dont la pression osmotique correspond à la pression physiologique pour l'organisme étudié.
Méiose	Mode de division propre aux lignées germinales des organismes recourant à la reproduction sexuée. Elle permet de produire les gamètes qui ne portent que la moitié du patrimoine génétique de l'individu, à partir d'une cellule mère, elle permet de donner 4 gamètes. La rencontre de deux gamètes, un mâle et un femelle, permettra de reconstruire un patrimoine génétique complet et un nouvel individu. La méiose joue un rôle important dans le brassage génétique.
Mésophile	Caractéristique d'un organisme pour lequel les températures moyennes (environ 35-40 °C) sont les plus favorables au développement.
Métabolomique	Etude du fonctionnement d'un système vivant par analyse de l'ensemble de son réseau métabolique, des processus de régulation et des interactions qui agissent dessus.

Métabolisme	Ensemble des réactions biochimiques qui permettent le fonctionnement d'une cellule. On distingue deux sous ensemble : le catabolisme qui regroupe les réactions de dégradation, et l'anabolisme qui regroupe toutes les réactions de fabrication.
Mitochondrie	Organite rencontré chez toutes les cellules eucaryotes, il est spécialisé dans la production de l'énergie, et particulièrement dans la respiration.
Mitose	Mode de division générale de toutes les cellules eucaryotes. Ce mode de division est une multiplication asexuée des cellules. Toutes les cellules que l'on obtient par mitoses sont des clones. A partir d'une cellule mère, on obtient 2 cellules filles. C'est ce mécanisme qui permet la croissance et la régénération d'un organisme.
Monotriche	Se dit d'une ciliature constituée d'un flagelle unique.
Neutrophile	Caractéristique d'un organisme dont le développement est le plus efficace à pH neutre (environ 7).
Nucléosome	Unité de base de la chromatine, constituée de deux tours de la séquence nucléotidique autour d'un groupe de 8 histones.
Nucléotide	Brique élémentaire des acides nucléiques, composée d'un phosphate, d'un sucre, et d'une base azotée. La base azotée, adénine, guanine, thymine, cytosine, uracile permet de désigner par sa première lettre le nucléotide. Le sucre sera du ribose ou du désoxyribose selon respectivement que l'acide nucléique est de l'ARN ou de l'ADN
Organelle	Anglicisme désignant les organistes. Ce terme sert pour certains à désigner spécifiquement les organites non compartimentés.
Organite	Nom désignant les structures spécialisées que l'on rencontre au sein d'une cellule. On distingue les organites compartimentés, délimités par une membrane, comme les mitochondries, réticulum endoplasmiques; chloroplastes, etc... des organistes non compartimentés comme les ribosomes, centrioles, etc...
Osmose	Diffusion de l'eau, de là où elle est le plus concentrée (où elle solubilise le moins de composés) vers là où elle est le moins concentrée (ou elle solubilise le plus de composés).
Péritriche	Se dit d'une ciliature constituée d'une multitude de flagelles répartis sur l'ensemble de la surface cellulaire et sans orientation privilégiée dans l'espace.
Perméase	Protéine assurant la fonction de transport à travers la membrane plasmique.
Phénotype	Ensemble des caractères observables d'un individu, c'est-à-dire l'ensemble des caractéristiques exprimées par le génotype.
Photosynthèse	Mécanisme biochimique complexe caractéristiques du règne végétal et de quelques bactéries au cours duquel l'énergie lumineuse est exploitée pour produire des sucres à partir de l'eau et du dioxyde de carbone.
Plasmolyse	Etat d'une cellule plongée dans un milieu hypertonique.

Porine	Protéine canal enchâssée dans une membrane, elle forme un "trou" dans la membrane qui facilite la diffusion.
Procaryote	Caractéristique des cellules dépourvues de "vrai" noyau.
Protiste	Règne assez mal défini du vivant, dont l'existence est souvent débattue, qui regroupe les microorganismes unicellulaires eucaryotes non autotrophes.
Psychrophile	Caractéristique d'un organisme pour lequel les températures basses (environ 10-15 °C) sont les plus favorables au développement.
Réticulum Endoplasmique	Organite rencontré chez toutes les cellules eucaryotes, il est spécialisé sans la maturation des protéines fraîchement synthétisées. Il est lié à l'enveloppe nucléaire et peut être rugueux ou lisse selon qu'il est associé à des ribosomes ou non.
Ribosome	Organite non compartimenté dont la fonction est la synthèse des protéines à partir des séquences d'ARN.
Scissiparité	Mécanisme de division des cellules procaryotes, il consiste en l'allongement de la cellule accompagné de la duplication du matériel génétique, suivi lorsqu'une taille critique est atteinte, de la séparation en deux cellules filles.
Sporophore	Organe reproducteur des eumycètes, il s'agit de la structure temporaire qui correspond à ce que l'on nomme couramment "champignon".
Symbiose	Association réciproquement profitable et durable entre des organismes appartenant à des espèces différentes.
Symport	Cotransport dans lequel les deux composés sont transportés dans le même sens.
Thermogénèse	Capacité à produire de la chaleur chez les animaux dits "à sang chaud".
Thermophile	Caractéristique d'un organisme pour lequel les températures moyennes (environ 55-65 °C) sont les plus favorables au développement.
Transcriptomique	Etude du fonctionnement d'un système vivant par analyse de l'ensemble de son profil d'expression génétique.
Vacuole	Organite dédié au stockage liquide, principalement chez les cellules végétales. Sa fonction est soit de stocker de l'eau, soit de stocker différentes molécules solubles avec un rôle de réserve ou de protection.
Vésicule	Organite dédié au stockage liquide chez les cellules animales. Sa fonction est soit de stocker de l'eau, soit de stocker différentes molécules solubles avec un rôle de réserve ou de protection.

